Electronic Journal of Probability

Electron. J. Probab. **27** (2022), article no. 150, 1–32.
ISSN: 1083-6489 https://doi.org/10.1214/22-EJP874

Quantitative mean-field limit for interacting branching diffusions*

Joaquín Fontbona† Felipe Muñoz-Hernández†‡

Abstract

We establish an explicit rate of convergence for some systems of mean-field interacting diffusions with logistic binary branching, towards solutions of nonlinear evolution equations with non-local self-diffusion and logistic mass growth, which were shown to describe their large population limits in [12]. The proof relies on a novel coupling argument for binary branching diffusions based on optimal transport, allowing us to sharply mimic the trajectory of the interacting binary branching population by means of a system of independent particles with suitably distributed random space-time births. We are thus able to derive an optimal convergence rate, in the dual bounded-Lipschitz distance on finite measures, for the empirical measure of the population, from the convergence rate in 2-Wasserstein distance of empirical distributions of i.i.d. samples. Our approach and results extend propagation of chaos techniques and ideas, from kinetic models to stochastic systems of interacting branching populations, and appear to be new in this setting, even in the simple case of pure binary branching diffusions.

Keywords: branching diffusions; population dynamics; mean-field limit; rate of convergence; optimal transport.
MSC2020 subject classifications: 92D25; 60J85; 60H30; 35Q92.
Submitted to EJP on September 27, 2021, final version accepted on October 31, 2022.
Supersedes arXiv:2101.04099.

1 Introduction and main result

Mathematical models of interacting and randomly evolving populations have been intensively studied the last decades through probabilistic and analytic approaches. Both viewpoints are able to integrate several biologically or ecologically meaningful features including: individuals' displacements, reproduction and deaths, competition

*J.F. acknowledges partial support from Fondecyt Grant 1201948 and from ANID-Chile BASAL Funds ACE210010 and FB21000 Center for Mathematical Modeling. F. M.-H. acknowledges financial support from Doctoral Fellowship ANID-PFCHA/Doctorado Nacional/2017-21171912.

†Universidad de Chile, Chile. E-mail: fontbona@dim.uchile.cl,fmunozh@dim.uchile.cl
‡Institut Polytechnique de Paris, France.

for resources, selection, and dispersive or attractive interactions. While PDE and analysis methods can provide aggregate deterministic descriptions of the collective or macroscopic behavior of such populations (see [16, 2, 3, 10, 11] and [6], to name but a few works), probabilistic methods have successfully been employed to describe the random behaviors and interactions of individuals at the microscopic (or finite population) level. Moreover, probabilistic weak convergence tools can be used to justify, in a rigorous way, how certain nonlinear evolution PDEs arise as limits in law of the empirical processes of individual-based models, when the population size goes to infinity (see for example [14, 1, 12, 5] and [4]). Nevertheless, although it is clear that certain law of large numbers for exchangeable random systems underlies the passage from the microscopic to the macroscopic scale in branching population models, the speed of this convergence is not explicitly known, even in the simple case of pure binary branching diffusions.

In this work, we develop a probabilistic approach to obtain quantitative convergence estimates for the large population limit of a general class of spatially branching diffusions with logistic growth and mean-field interactive spatial dynamics. The population and its evolution are described by a right-continuous measure-valued Markov process taking values for fixed $K \in \mathbb{N} \setminus \{0\}$ in the space of weighted finite point measures over $\mathbb{R}^d$

$$\mathcal{M}^K(\mathbb{R}^d) \coloneqq \left\{ \frac{1}{K} \sum_{n=1}^{N} \delta_{x^n} : x^n \in \mathbb{R}^d, N \in \mathbb{N} \setminus \{0\} \right\} \subseteq \mathcal{M}^+(\mathbb{R}^d).$$

Here, $\mathcal{M}^+(\mathbb{R}^d)$ stands for the space of finite nonnegative measures on $\mathbb{R}^d$ endowed with the weak topology and δ_x is the Dirac mass at $x \in \mathbb{R}^d$. We denote said process by

$$\mu_t^K = \frac{1}{K} \sum_{n=1}^{N_t^K} \delta_{X_t^{n,K}}, \quad t \geq 0,$$

where $N_t^K \coloneqq K\langle \mu_t^K, 1\rangle \in \mathbb{N}$, with $\langle \cdot, \cdot \rangle$ denoting the integral pairing, is the number of living individuals at time $t \geq 0$ and $X_t^{1,K}, \dots, X_t^{N_t^K,K}$ are their positions in $\mathbb{R}^d$. The parameter K measures the population size and can be interpreted as the carrying capacity of the underlying environment (see [1]).

The dynamics of $(\mu_t^K)_{t\geq 0}$ is summarized as follows:

- The initial population is described by a random measure $\mu_0^K \in \mathcal{M}^K(\mathbb{R}^d)$.
- Each living individual carries at each instant $t > 0$ two clocks independent between them: one reproduction clock, exponential of parameter $r > 0$ and independent of everything else in the system, and one mortality clock, conditionally exponential of parameter cN_t^K/K, for $c > 0$, given the population size N_t^K. If the reproduction clock of a particle rings at time t when at position x, it gives birth to a new particle at that same position. If the mortality clock rings the particle disappears. Equivalently, the process jumps from μ_{t-}^K to $\mu_t^K = \mu_{t-}^K + K^{-1}\delta_x$ in the first case and to $\mu_t^K = \mu_{t-}^K - K^{-1}\delta_x$ in the second.
- Between birth or death events, for each $n = 1, ..., N_t^K$, the individual $X_t^{n,K}$ evolves according to the diffusion process

$$\mathrm{d}X_t^{n,K} = b\big(X_t^{n,K}, H * \mu_t^K(X_t^{n,K})\big)\,\mathrm{d}t + \sigma\big(X_t^{n,K}, G * \mu_t^K(X_t^{n,K})\big)\,\mathrm{d}B_t^n,$$

  where $(B^n)_{n\geq 1}$ are Brownian motions in $\mathbb{R}^d$, independent between them and independent of μ_0^K and of the birth and death clocks. In particular, the drift and the diffusion coefficients are affected by the local concentration of individuals, through the convolution of the empirical measure μ_t^K with the kernels H and G, respectively.

This model is a subclass of the non-local Lotka-Volterra cross-diffusion systems introduced in [12] as a microscopic, individual-based counterpart of the celebrated Shigesada-Kawasaki-Teramoto cross-diffusion system [16]. Here, we consider a simplified setting, consisting in one single species with self-interaction at the individuals' displacements level, and we assume that the demographic parameters determining births and deaths are spatially homogeneous. In particular, the competitive pressure exerted on each individual is global and proportional to the population size, which corresponds to a constant competition kernel in the general model of [12].

Following [12] one can prove that, when K goes to infinity, for each $T>0$ the empirical measure process $(\mu^K_t)_{t\in[0,T]}$ converges in law (in the Skorokhod space of finite measure-valued paths on $[0,T]$) to a deterministic continuous measure-valued function $(\mu_t)_{t\in[0,T]}$, which is the unique weak solution of a non-local self-diffusion equation (see (1.2) below). The following are assumptions that ensure this convergence and which will be required to establish our main result.

Hypothesis (H):

H.1. $\mu_0\in\mathcal{M}^+(\mathbb{R}^d)$ is a given non-null measure and the random measures $(\mu^K_0)_{K\in\mathbb{N}\setminus\{0\}}\subseteq\mathcal{M}^K(\mathbb{R}^d)$ are such that the sequence $(\langle\mu^K_0,1\rangle)_{K\in\mathbb{N}\setminus\{0\}}$ converges in law as $K\to\infty$ to $\langle\mu_0,1\rangle$. Moreover, for each $K\in\mathbb{N}\setminus\{0\}$, conditionally on $\langle\mu^K_0,1\rangle$, the measure μ^K_0 is supported on $N^K_0=K\langle\mu^K_0,1\rangle$ i.i.d. random variables with common law not depending on K and given by the normalized measure $\bar{\mu}_0:=\mu_0/\langle\mu_0,1\rangle$.

H.2. The functions $\sigma\colon\mathbb{R}^d\times\mathbb{R}_+\to\mathbb{R}^{d\otimes d}$ and $b\colon\mathbb{R}^d\times\mathbb{R}_+\to\mathbb{R}^d$ (with $\mathbb{R}^{d\otimes d}$ denoting the space of $d\otimes d$ matrices) are Lipschitz continuous. Moreover, there exists $C_\sigma>0$ such that

$$|\sigma(x,v)|\le C_\sigma(1+v),\quad\forall x\in\mathbb{R}^d,v\in\mathbb{R}_+.$$

H.3. The functions $G,H\colon\mathbb{R}^d\to\mathbb{R}_+$ are bounded and Lipschitz continuous.

Remark 1. Given $\mu_0\in\mathcal{M}^+(\mathbb{R}^d)$, it is always possible to construct a sequence of random measures $(\mu^K_0)_{K\in\mathbb{N}\setminus\{0\}}\subseteq\mathcal{M}^K(\mathbb{R}^d)$, such that (H.1) holds. See Lemma 9 c) for an example of such construction.

Under assumption (H), $(\mu^K_t)_{t\ge0}$ is a Markov process which has finitely many jumps in each finite time interval and whose law is uniquely determined. See [12] for details and [9] for general background on measure-valued Markov processes.

Let $a:=\sigma\sigma^{\mathrm{t}}$ and, given $\mu\in\mathcal{M}^+(\mathbb{R}^d)$, define an operator acting on $C^2(\mathbb{R}^d)$ functions ϕ by

$$L_\mu\phi(x)=\frac{1}{2}\mathrm{Tr}\left(a(x,G*\mu(x))\mathrm{Hess}(\phi)(x)\right)+b(x,H*\mu(x))\cdot\nabla\phi(x).\tag{1.1}$$

Then, as a particular case of [12, Theorem 3.1], we have the following statement.

Theorem 2. Assume (H) with $\mu_0\in\mathcal{M}^+(\mathbb{R}^d)$ given and that $\sup_{K\in\mathbb{N}\setminus\{0\}}\mathbb{E}(\langle\mu^K_0,1\rangle^p)<\infty$ for some $p\ge3$. Then, as $K\to\infty$, the sequence of processes $(\mu^K)_{K\in\mathbb{N}\setminus\{0\}}$ converges in law in $D([0,T],\mathcal{M}^+(\mathbb{R}^d))$ to the unique (deterministic) continuous finite measure-valued function $(\mu_t)_{t\in[0,T]}$ solution of

$$\langle\mu_t,f(t,\cdot)\rangle=\langle\mu_0,f(0,\cdot)\rangle+\int_0^t\left\langle\mu_s,\partial_sf(s,\cdot)+L_{\mu_s}f(s,\cdot)+(r-c\langle\mu_s,1\rangle)f(s,\cdot)\right\rangle\mathrm{d}s,\quad\forall t\in[0,T],\tag{1.2}$$

for every $f\in C^{1,2}_b([0,T]\times\mathbb{R}^d)$ such that $\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}(1+|x|)|\nabla f(t,x)|<\infty$.

Notice that the total mass $n_t:=\langle\mu_t,1\rangle$ of the measure μ_t has an autonomous, logistic evolution in $(0,\infty)$, that is

$$\partial_tn_t=(r-cn_t)n_t,\quad t\ge0.$$

Our main result is the quantification of the convergence to the large population limit in Theorem 2. Recall that the weak topology on the space $\mathcal{M}^+(\mathbb{R}^d)$ can be metrized by means of the dual bounded-Lipschitz norm, which we denote by $\|\cdot\|_{\mathrm{BL}^*}$ (see Section 2 for details). We have:

Theorem 3. Assume (H) with $\mu_0 \in \mathcal{M}^+(\mathbb{R}^d)$ given, that $\sup_{K\in\mathbb{N}\setminus\{0\}} \mathbb{E}(\langle \mu_0^K, 1\rangle^p) < \infty$ for some $p \geq 4$, and that $\int_{\mathbb{R}^d} |x|^q \,\mu_0(\mathrm{d}x) < \infty$ for some $q > 2$. Then, for all $K \in \mathbb{N}\setminus\{0\}$ and $T > 0$ one has

$$\sup_{t\in[0,T]} \mathbb{E}\Big(\big\|\mu_t^K - \mu_t\big\|_{\mathrm{BL}^*}\Big) \leq C_T\Big(I_4(K) + \sqrt{R_{d,q}(K)}\Big),$$

where $C_T > 0$ depends on T, p, q, and on the data of the model, $I_4(K) = \mathbb{E}\big(|\langle \mu_0^K, 1\rangle - \langle \mu_0, 1\rangle|^4\big)^{\frac{1}{4}}$, and $R_{d,q}\colon \mathbb{N}\setminus\{0\} \to \mathbb{R}_+$ is the function defined by

$$R_{d,q}(K) \coloneqq K^{-\frac{(q-2)}{q}} + \begin{cases} K^{-\frac{1}{2}}, & \text{if } d < 4 \text{ and } q \neq 4,\\ K^{-\frac{1}{2}}\log(1+K), & \text{if } d = 4 \text{ and } q \neq 4,\\ K^{-\frac{2}{d}}, & \text{if } d > 4 \text{ and } q \neq \frac{d}{d-2}.\end{cases} \tag{1.3}$$

The fact that $\langle \mu_0^K, 1\rangle$ converges at least as fast as $K^{-1/4}$ in L^4 to $\langle \mu_0, 1\rangle$ can be granted for large families of random measures satisfying (H.1) (see Lemma 9 in Section 2.2 for details as well as for possible relaxations of assumption (H.1)). The convergence rate in Theorem 3 thus essentially depends non-increasingly on the dimension d, and on the amount of finite moments of the measure μ_0. For modeling purposes, the most relevant setting is $d = 3$, in which case the rate is equivalent to $K^{-1/4}$ if $q \in [4, +\infty)$, or to the slower rate $K^{-(q-2)/(2q)}$ if $q \in (2, 4)$. We notice also that the same result can be obtained in the case that each individual of the population additionally carries an independent, autonomous exponential killing clock of a given fixed parameter (with the natural modification of the limiting PDE).

To prove Theorem 3 we will extend to the branching populations setting some probabilistic coupling techniques based on optimal transport, which were developed to quantify propagation of chaos in binary interacting particle systems from kinetic theory [7, 8]. See [17] and [15] for general background on propagation of chaos theory.

In the next section, we establish some preliminary results and present the strategy of the proof of Theorem 3, along with an outline of the remainder of the paper. We shall also discuss the ideas underlying our approach and some consequences of our main result, in the light of propagation of chaos theory.

2 Preliminaries and strategy of the proof

Denote by $\mathrm{BL}(\mathbb{R}^d)$ the space of real Lipschitz continuous bounded functions in $\mathbb{R}^d$ with the norm

$$\|\varphi\|_{\mathrm{BL}} = \sup_{x\neq y} \frac{|\varphi(x) - \varphi(y)|}{x - y} + \sup_x |\varphi(x)|,$$

and by $\|\cdot\|_{\mathrm{BL}^*}$ the corresponding dual norm on the space $\mathcal{M}(\mathbb{R}^d)$ of finite signed measures on $\mathbb{R}^d$. The induced distance

$$\|\mu - \nu\|_{\mathrm{BL}^*} = \sup_{\|\varphi\|_{\mathrm{BL}}\leq 1} |\langle \mu - \nu, \varphi\rangle|,$$

is well known to generate the weak convergence topology on $\mathcal{M}^+(\mathbb{R}^d)$. The subspace of $\mathcal{M}^+(\mathbb{R}^d)$ of probability measures is denoted by $\mathcal{P}(\mathbb{R}^d)$. Given a measure $\mu \in \mathcal{M}^+(\mathbb{R}^d)$, its q-th moment for $q \in [1, \infty)$ is denoted by

$$M_q(\mu) = \int_{\mathbb{R}^d} |x|^q \,\mu(\mathrm{d}x). \tag{2.1}$$

For $p \in [1, \infty)$, the p-Wasserstein distance $W_p(\mu, \nu)$ between two probability measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ is defined by

$$W_p(\mu, \nu) = \left(\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p \, \pi(\mathrm{d}x, \mathrm{d}y) \right)^{\frac{1}{p}},$$

where $\Pi(\mu, \nu)$ is the set of probability measures over $\mathbb{R}^d \times \mathbb{R}^d$ that have μ and ν as first and second marginals respectively. A coupling $\pi \in \Pi(\mu, \nu)$ realizing the infimum always exists and is called an *optimal coupling* between μ and ν for the transport cost $c(x, y) = |x - y|^p$. W_p defines a complete distance if restricted to the space $\{\mu \in \mathcal{P}(\mathbb{R}^d) : M_p(\mu) < \infty\}$ and is equivalent therein to the weak topology strengthened with the convergence of p-th moments. See [18] for background.

For every $\mu \in \mathcal{M}^+(\mathbb{R}^d)$, we will throughout denote by $\bar{\mu}$ the probability measure on $\mathbb{R}^d$ obtained from it by normalization:

$$\bar{\mu} \coloneqq \frac{1}{\langle \mu, 1 \rangle} \mu \in \mathcal{P}(\mathbb{R}^d). \tag{2.2}$$

This notation is consistent with the relation between μ_0 and $\bar{\mu}_0$ in assumption (H.1).

The following simple relations for finite measures, proved in Appendix, will be useful.

Lemma 4. Let $\mu, \nu \in \mathcal{M}^+(\mathbb{R}^d)$ and let $\bar{\mu}, \bar{\nu} \in \mathcal{P}(\mathbb{R}^d)$ be their corresponding normalized versions defined as in (2.2). We have

$$\|\mu - \nu\|_{\mathrm{BL}^*} \le \langle \mu, 1 \rangle \|\bar{\mu} - \bar{\nu}\|_{\mathrm{BL}^*} + \big| \langle \mu, 1 \rangle - \langle \nu, 1 \rangle \big|,$$

and

$$\|\bar{\mu} - \bar{\nu}\|_{\mathrm{BL}^*} \le \inf_{\pi \in \Pi(\bar{\mu}, \bar{\nu})} \int |x - y| \wedge 2 \, \pi(\mathrm{d}x, \mathrm{d}y) \le W_1(\bar{\mu}, \bar{\nu}).$$

The basic estimate on which our main result relies, is the quantitative bound in 2-Wasserstein distance for empirical measures of i.i.d. samples, proved in [13] and stated next for convenience.

Theorem 5. Let $\tilde{\mu} \in \mathcal{P}(\mathbb{R}^d)$ be given and $(X^n)_{n \in \mathbb{N}}$ be i.i.d. random variables with law $\tilde{\mu}$. Assume that $M_q(\tilde{\mu}) < \infty$ for some $q > 2$, with $M_q(\tilde{\mu})$ defined as in (2.1). Then, there exists a constant $C_{d,q} > 0$ depending only on d and q such that, for all $N \in \mathbb{N} \setminus \{0\}$,

$$\mathbb{E}\left(W_2^2 \left(\frac{1}{N} \sum_{n=1}^{N} \delta_{X^n}, \tilde{\mu} \right) \right) \le C_{d,q} M_q^{\frac{2}{q}}(\tilde{\mu}) \, R_{d,q}(N),$$

with $R_{d,q}$ defined as in (1.3).

We can deduce analogous estimates for random empirical measures in $\mathcal{M}^K(\mathbb{R}^d)$ with atoms satisfying a certain conditional independence property:

Lemma 6. Let $\mu \in \mathcal{M}^+(\mathbb{R}^d)$ be such that $M_q(\mu) < \infty$ for some $q > 2$, with $M_q(\mu)$ defined in (2.1). Let also N be a random variable in $\mathbb{N}$ with $\mathbb{E}(N) < \infty$ and ν^K a random variable in $\mathcal{M}^K(\mathbb{R}^d)$ that, conditionally on N, is supported on N atoms that are i.i.d. random variables of law $\bar{\mu}$ defined from μ as in (2.2). Then, there exists a constant $C_{d,q} > 0$ that depends only on d and q such that

$$\mathbb{E}\Big(\frac{N}{K} W_2^2 \big(\bar{\nu}^K, \bar{\mu} \big) \Big) \le C_{d,q} M_q^{\frac{2}{q}}(\bar{\mu}) \, \mathbb{E}(1 \vee (N/K)) R_{d,q}(K),$$

where $\bar{\nu}^K$ is the normalized version of ν^K as in (2.2) and with $R_{d,q}$ defined as in (1.3).

See Appendix for the proof. Notice that under assumption (H.1), Lemma 6 immediately provides quantitative estimates for $W_2^2(\bar{\mu}_t^K, \bar{\mu}_t)$ when $t = 0$; however, the required conditional independence property is lost as soon as $t > 0$, even in the case of pure branching diffusions.

2.1 Proof strategy and plan of the paper

The proof of Theorem 3 is based on the construction, for each K, of a coupling between the system $(\mu_t^K)_{t\geq 0}$, and an auxiliary system of particles in $\mathcal{M}^K(\mathbb{R}^d)$ denoted by

$$\nu_t^K := \frac{1}{K}\sum_{n=1}^{N_t^K} \delta_{Y_t^{n,K}}, \quad t \geq 0,$$

such that the following condition holds:

Condition (C):

C.1. $\nu_0^K = \mu_0^K$ and $K\langle \nu_t^K, 1\rangle = K\langle \mu_t^K, 1\rangle = N_t^K$ for all $t \geq 0$ almost surely.

C.2. For each $t \geq 0$, conditionally on $\langle \nu_t^K, 1\rangle$, the atoms of ν_t^K are i.i.d. random variables of law $\bar{\mu}_t = \mu_t / \langle \mu_t, 1\rangle$.

C.3. For each $T > 0$ there is a constant $C_T > 0$ depending on T and on the data of Theorem 3 such that

$$\mathbb{E}\Big(\frac{N_t^K}{K} W_2^2\big(\bar{\nu}_t^K, \bar{\mu}_t^K\big)\Big) \leq C_T\,(R_{d,q}(K) + I_4^2(K)), \quad \forall t \in [0, T],$$

with $I_4(K)$ defined in the statement of Theorem 3.

Let us describe how this construction will be used and how the arguments of the proof will unfold in the remainder of the paper:

- Thanks to condition (C.1), Lemma 4 and some auxiliary estimates, obtaining the desired bound boils down, by triangular inequality, to controlling uniformly on $t \in [0, T]$ for each $T > 0$, the quantities $\mathbb{E}\big(\frac{N_t^K}{K} W_2^2\big(\bar{\nu}_t^K, \bar{\mu}_t^K\big)\big)$ and $\mathbb{E}\big(\frac{N_t^K}{K} W_2^2\big(\bar{\nu}_t^K, \bar{\mu}_t\big)\big)$.
- Condition (C.2) and Lemma 6 together imply that the quantity $\mathbb{E}\big(\frac{N_t^K}{K} W_2^2\big(\bar{\nu}_t^K, \bar{\mu}_t\big)\big)$ is bounded by $C_T\, R_{d,q}(K)$ for all $t \in [0, T]$.
- The previous facts and the bound in condition (C.3), together, will imply the bound asserted in Theorem 3.

In Section 3 we explicitly construct the coupled particle systems, $(\mu_t^K)_{t\geq 0}$ and $(\nu_t^K)_{t\geq 0}$, using common Brownian motions and a suitable Poisson point measure. In this construction, condition (C.1) will be simply verified since the two systems will start from the same state and their corresponding birth and death events will be simultaneous. In order to ensure condition (C.2), each atom $Y_t^{n,K}$ of ν_t^K will be defined as a suitable McKean-Vlasov diffusion (defined in Proposition 11), which will evolve independently of everything else in the system after being born and will have the law $\bar{\mu}_t$ at each time t from that moment on.

The crucial, far from obvious feature of the coupling is condition (C.3). Since we have

$$\mathbb{E}\Big(\frac{N_t^K}{K} W_2^2\big(\bar{\nu}_t^K, \bar{\mu}_t^K\big)\Big) \leq \mathbb{E}\bigg(\frac{N_t^K}{K}\frac{1}{N_t^K}\sum_{n=1}^{N_t^K} \|X_t^{n,K} - Y_t^{n,K}\|^2\bigg) = \mathbb{E}\bigg(\frac{1}{K}\sum_{n=1}^{N_t^K} \|X_t^{n,K} - Y_t^{n,K}\|^2\bigg),$$

and the coefficients are Lipschitz, coupling particles $(X_t^{n,K}, Y_t^{n,K})$ by using the same driving Brownian motion will allow us to keep their trajectories close and hence $W_2^2\big(\bar{\nu}_t^K, \bar{\mu}_t^K\big)$ small in between birth or death events. However, ensuring (C.3) will moreover require that the birth positions of the two particles be coupled too, in the best possible way (in the L^2 sense). This is where optimal transport ideas and techniques introduced in [7, 8] will come into play. Indeed, on one hand, the birth position of a new particle in the system $(\nu_t^K)_{t\geq 0}$, born at a random time s, is sampled in $\mathbb{R}^d$ according to the law $\bar{\mu}_s$.

On the other, choosing randomly a particle that branches at time s in system $(\mu_t^K)_{t\geq 0}$ is equivalent to sampling a position in $\mathbb{R}^d$ according to the empirical law $\bar{\mu}_{s-}^K$. Thus, the optimal way to couple a pair of atoms in the two systems at their birth time s is to sample them simultaneously from the optimal coupling for W_2^2 of the law $\bar{\mu}_s$ and the (random) law $\bar{\mu}_{s-}^K$. This joint sampling must be done in a measurable way in terms of the state of the process at time $s-$, which requires adapting a non-trivial construction from [7], which we do in Lemma 13.

In Section 4 we consider the simpler case of pure binary branching processes (i.e. with no mean-field interaction between the particles nor competition). We establish some auxiliary estimates, we prove that condition (C.3) holds in that particular case, and we deduce Theorem 3 in this specific setting, with slightly better bounds. In Section 5 we follow similar steps to deduce the proof of Theorem 3 as stated in the general case. Finally, in the last section, Section 6, we discuss potential extensions of the developed ideas and results to more general branching population models.

Before delving into the proofs, we briefly discuss the relation of our results with the propagation of chaos property, and comment on condition (H.1) and extensions of it in that framework.

2.2 Propagation of chaos for interacting branching diffusions

It is well known that convergence of the empirical probability distribution of N exchangeable particles to some deterministic probability measure, when N is a non-random integer that goes to infinity, is equivalent to the property of propagation of chaos, or asymptotic independence of the particles [17, 15]. The following generalization allows us to see Theorem 3 as a propagation of chaos type result.

Definition 7. Let $(N^K)_{K\in\mathbb{N}\setminus\{0\}}$ be random variables in $\mathbb{N}$ going in law to ∞ as $K\to\infty$. We say that a family $((Y^{1,K},\dots,Y^{N^K,K}))_{K\in\mathbb{N}\setminus\{0\}}$ of random vectors, $(\mathbb{R}^d)^{N^K}$-valued and exchangeable conditionally on N^K for each K, is conditionally P-*chaotic* given $(N^K)_{K\in\mathbb{N}\setminus\{0\}}$ if for some $P\in\mathcal{P}(\mathbb{R}^d)$ and every $j\in\mathbb{N}\setminus\{0\}$ the (random) conditional distributions $\left(\mathcal{L}(Y^{1,K},\dots,Y^{j\wedge N^K,K}\mid N^K)\right)_{K\in\mathbb{N}\setminus\{0\}}$ given N^K and the event $\{N^K\geq j\}$, converge in law in $\mathcal{P}((\mathbb{R}^d)^j)$ to $P^{\otimes j}$ as $K\to\infty$.

In the case that $N^K=K$ for all $K\in\mathbb{N}\setminus\{0\}$, one recovers the well known notion of P-chaoticity [17, 15]. We deduce the following result, proved in Section 5.

Corollary 8. Under the same assumptions of Theorem 3, we have that for each $t\geq 0$ the family $((X_t^{1,K},\dots,X_t^{N_t^K,K}))_{K\in\mathbb{N}\setminus\{0\}}$ is conditionally P_t-chaotic given $(N_t^K)_{K\in\mathbb{N}\setminus\{0\}}$, with $P_t=\mu_t/\langle\mu_t,1\rangle$.

We end this section gathering some remarks about assumption (H.1), including its possible relaxation to a chaoticity condition. The proof of this result is given in the Appendix.

Lemma 9. a) Under (H.1), $(\mu_0^K)_{K\in\mathbb{N}\setminus\{0\}}$ converges in law as $K\to\infty$ to $\mu_0\in\mathcal{M}^+(\mathbb{R}^d)$.

b) Let $\tilde{\mu}_0\in\mathcal{P}(\mathbb{R}^d)$ be given and for each $K\in\mathbb{N}\setminus\{0\}$ let $\mu_0^K\in\mathcal{M}^K(\mathbb{R}^d)$ be a random point measure. Assume that $(\langle\mu_0^K,1\rangle)_{K\in\mathbb{N}\setminus\{0\}}$ converges in law as $K\to\infty$ to a constant in $(0,\infty)$ and that there exists a $\tilde{\mu}_0$-chaotic family of exchangeable random vectors

$$((Y^{1,N},\dots,Y^{N,N}):N\in\mathbb{N}\setminus\{0\}),$$

such that for all K, conditionally on $K\langle\mu_0^K,1\rangle=N$, the set of atoms of μ_0^K has the same law as $(Y^{1,N},\dots,Y^{N,N})$. Then, $(\mu_0^K)_{K\in\mathbb{N}\setminus\{0\}}$ converges in law as $K\to\infty$ to a (deterministic) limit $\mu_0\in\mathcal{M}^+(\mathbb{R}^d)$ now given by

$$\mu_0 \coloneqq \lim_{K\to\infty}\langle\mu_0^K,1\rangle\tilde{\mu}_0. \tag{2.3}$$

and (with the notation (2.2)) one has $\bar{\mu}_0 = \tilde{\mu}_0$.

c) Given $\mu_0 \in \mathcal{M}^+(\mathbb{R}^d)$, assumption (H.1) holds if, for example, for each K we choose $K\mu_0^K$ to be a Poisson point measure on $\mathbb{R}^d$ of intensity $K\mu_0$. In this case $I_4(K)$ given in Theorem 3 satisfies $I_4(K) \leq CK^{-1/2}$.

Remark 10. If instead of (H.1) we only suppose that $(\mu_0^K)_{K\in\mathbb{N}\setminus\{0\}}$ satisfies the condition given in Lemma 9 b), Theorem 2 still holds with μ_0 given by (2.3), and Theorem 3 holds but with an additional term on the r.h.s. of generic form: $C_T\mathbb{E}\big(\frac{1}{K}\sum_{n=1}^{N_0^K}\big\|X_0^{n,K} - Y_0^{n,K}\big\|^2\big)$, where $\big((X_0^{1,K},\ldots,X_0^{N,K}),(Y_0^{1,K},\ldots,Y_0^{N,K})\big)$ is for each $N,K \in \mathbb{N}$, conditionally on $\{N_0^K = N\}$, a coupling of the N atoms of μ_0^K and an i.i.d. sample of size N of the law $\tilde{\mu}_0$. See Remark 27 for details and for the optimal value of this term.

3 Pathwise constructions and coupling algorithm

For the rest of the article we will omit the superscripts K in the particles' positions, e.g. we write $\big(X_t^1,\ldots,X_t^{N_t^K}\big) = \big(X_t^{1,K},\ldots,X_t^{N_t^K,K}\big)$ since we will be working with fixed $K \in \mathbb{N}\setminus\{0\}$ and no ambiguity is possible.

We will construct both systems $(\mu_t^K = \frac{1}{K}\sum_{n=1}^{N_t^K}\delta_{X_t^n})_{t\geq 0}$ and $(\nu_t^K = \frac{1}{K}\sum_{n=1}^{N_t^K}\delta_{Y_t^n})_{t\geq 0}$ from the following set of independent stochastic inputs, defined in a common complete probability space $(\Omega,\mathcal{F},\mathbb{P})$:

- A sequence $(W^j)_{j\geq 1}$ of independent Brownian motions in $\mathbb{R}^d$.
- A Poisson point measure $\mathcal{N}(\mathrm{d}s,\mathrm{d}\rho,d\theta)$ on $[0,\infty)\times[0,\infty)\times[0,\infty)$, with intensity $\mathrm{d}s\otimes\mathrm{d}\rho\otimes\mathrm{d}\theta$.
- A sequence $(Z_0^j)_{j\geq 1}$ of i.i.d. random vectors of law $\bar{\mu}_0$, defined from μ_0 as in (2.2).
- A random variable N_0^K in $\mathbb{N}$.

We will also make use of a special diffusion process considered in [12], which can be seen as a nonlinear process in the sense of McKean [17, 15]. In the current setting, this process is characterized in the next result:

Proposition 11. Assume (H) and let $(\mu_t)_{t\geq 0}$ be the unique weak solution in $\mathcal{M}^+(\mathbb{R}^d)$ given by Theorem 2 of the nonlinear equation (1.2) with initial condition μ_0, written for short as

$$\frac{\partial\mu_t}{\partial t} = L_{\mu_t}^*\mu_t + \big(r - c\langle\mu_t,1\rangle\big)\mu_t, \quad t\geq 0,$$

with L_μ^* denoting the adjoint operator of L_μ in (1.1). Let W be a $d-$dimensional Brownian motion and Y_0 an independent random variable in $\mathbb{R}^d$ with law $\bar{\mu}_0 = \mu_0/\langle\mu_0,1\rangle$. Then, there is pathwise existence and uniqueness for the SDE

$$Y_t = Y_0 + \int_0^t b(Y_s, H*\mu_s(Y_s))\,\mathrm{d}s + \int_0^t \sigma(Y_s, G*\mu_s(Y_s))\,\mathrm{d}W_s, \quad t\geq 0. \tag{3.1}$$

Moreover, the flow of time-marginal laws of $(Y_t)_{t\geq 0}$ is the unique weak solution $(\tilde{\mu}_t)_{t\geq 0}$ in $\mathcal{P}(\mathbb{R}^d)$ of the linear, non-homogeneous in time Fokker-Planck equation

$$\frac{\partial\tilde{\mu}_t}{\partial t} = L_{\mu_t}^*\tilde{\mu}_t, \quad t\geq 0, \tag{3.2}$$

(w.r.t. test functions as in Theorem 2) with initial condition $\tilde{\mu}_0 = \bar{\mu}_0$, and for all $t\geq 0$ we have $\tilde{\mu}_t = \bar{\mu}_t$, the normalized version of μ_t (see (2.2)). Last, for every bounded measurable function $f\colon\mathbb{R}^d\to\mathbb{R}$ we have $\langle\mu_t,f\rangle = \mathbb{E}(f(Y_t)n_t)$, where n_t is the unique solution with $n_0 = \langle\mu_0,1\rangle$ of the logistic equation

$$\mathrm{d}n_t = \big(r - cn_t\big)n_t\,\mathrm{d}t. \tag{3.3}$$

The proof of Proposition 11 is postponed to Section 5.

Remark 12. a) The pathwise properties of the SDE (3.1) stated in Proposition 11 imply for fixed $\tau > 0$ that if Y'_τ is a random variable of law $\bar{\mu}_\tau$ independent of W, then the solution $(Y'_t)_{t\geq\tau}$ of the SDE

$$Y'_t = Y'_\tau + \int_\tau^t b(Y'_s, H * \mu_s(Y'_s))\,\mathrm{d}s + \int_\tau^t \sigma(Y'_s, G * \mu_s(Y'_s))\,\mathrm{d}W_s, \quad t \geq \tau,$$

has the same law as $(Y_t)_{t\geq\tau}$. In particular, Y'_t has law $\bar{\mu}_t$ for all $t \geq \tau$.

b) When σ and b in (H.2.) depend only on the position x and not on the nonnegative real variable v, the process (3.1) is the standard diffusion associated with the generator acting on $C^2(\mathbb{R}^d)$ functions ϕ:

$$L\phi(x) = \frac{1}{2}\mathrm{Tr}\big(a(x)\mathrm{Hess}\phi(x)\big) + b(x)\cdot\nabla\phi(x),$$

which in that case also drives each of the particles of the branching system $(\mu_t^K)_{t\geq 0}$. Notice also that in this setting, thanks to the Lipschitz character of the coefficients, if $\bar{\mu}_0$ has finite moments of order $q \geq 2$, then finiteness of these moments is uniformly propagated over any time interval $[0,T]$.

Last, the following construction, based on optimal transport and adapted from [7], will allow us to couple the births positions in the two systems in the most efficient way, as discussed in Section 2.1.

Lemma 13. Let $\mathbf{i}\colon \mathbb{R} \to \mathbb{N}$ denote the function defined by

$$\rho \mapsto \mathbf{i}(\rho) = \lfloor \rho \rfloor + 1,$$

and N be a positive integer. Let also $(\tilde{\mu}_t)_{t\geq 0}$ be a given weakly continuous flow of probability measures on $\mathbb{R}^d$ with finite second order moments. There exists a measurable mapping

$$\Lambda^N\colon \mathbb{R}_+ \times (\mathbb{R}^d)^N \times [0,N) \to \mathbb{R}^d, \quad (t,\mathbf{x},\rho) \mapsto \Lambda_t^N(\mathbf{x},\rho),$$

with the following properties:

- For every $t \geq 0$ and $\mathbf{x} = (x^1,\ldots,x^N) \in (\mathbb{R}^d)^N$, if ρ is uniformly chosen from $[0,N)$, then the pair $(\Lambda_t^N(\mathbf{x},\rho), x^{\mathbf{i}(\rho)})$ is an optimal coupling between $\tilde{\mu}_t$ and $\frac{1}{N}\sum_{i=1}^N \delta_{x^i}$ with respect to the cost function $(u,v)\mapsto |u-v|^2$.
- If $\mathbf{Y}$ is any exchangeable random vector in $(\mathbb{R}^d)^N$, then $\mathbb{E}\big(\int_{j-1}^j \phi(\Lambda_t^N(\mathbf{Y},\tau))\mathrm{d}\tau\big) = \langle \tilde{\mu}_t, \phi\rangle$ for any $j \in \{1,\ldots,N\}$, and any bounded measurable function ϕ.
- The function $\Lambda\colon \mathbb{N}\times\mathbb{R}_+\times\big(\bigcup_{N\in\mathbb{N}\setminus\{0\}}(\mathbb{R}^d)^N\big)\times\mathbb{R}_+ \to \mathbb{R}^d$ given by

$$\Lambda(N,t,\mathbf{x},\rho) = \Lambda_t^N\big((x^n)_{n=1}^N, \rho\wedge N\big),$$

if $\mathbf{x} = (x^n)_{n=1}^N \in (\mathbb{R}^d)^N$, and $0 \in \mathbb{R}^d$ otherwise, is measurable.

Proof. The proof of the first and second assertions is essentially the same as in [7, Lemma 3] (the only difference being that we do not remove one particle from the vector $\mathbf{x}$ to construct an empirical measure). The last assertion follows noting that $\Lambda^{-1}(A) = \bigcup_{N\neq 0}\{N\}\times(\Lambda^N)^{-1}(A)$ is a measurable set for any Borel set $A \in \mathbb{R}^d$ such that $0 \notin A$, and $\Lambda^{-1}(\{0\}) = \Big(\bigcup_{N\neq 0}\{N\}\times\mathbb{R}_+\times\cup_{n\neq N}(\mathbb{R}^d)^n\times\mathbb{R}_+\Big)\cup\Big(\bigcup_{N\neq 0}\{N\}\times(\Lambda^N)^{-1}(\{0\})\Big)$. □

3.1 Coupling algorithm

We now give an algorithm to construct $(\mu_t^K = \frac{1}{K}\sum_{n=1}^{N_t^K}\delta_{X_t^n})_{t\geq 0}$ and $(\nu_t^K = \frac{1}{K}\sum_{n=1}^{N_t^K}\delta_{Y_t^n})_{t\geq 0}$, jointly in the same probability space. Before doing so, we also introduce a sequence of *labelling processes*

$$(j_t(n) : t \geq 0)_{n\geq 1},$$

taking values in the positive integers, that will be dynamically defined to select from $(W^j)_{j\geq 1}$ the Brownian motions driving each coupled pairs of particles (X_t^n, Y_t^n), between reproduction or death events.

In the algorithm and in the remainder of this Section, $(\mu_t)_{t\geq 0}$ denotes the unique weak solution in $\mathcal{M}^+(\mathbb{R}^d)$ of the nonlinear equation (1.2) given by Theorem 2, and Λ stands for the function constructed in Lemma 13 with $(\tilde{\mu}_t)_{t\geq 0} = (\bar{\mu}_t)_{t\geq 0}$, the normalized version of $(\mu_t)_{t\geq 0}$.

Algorithm (A):

0. We set $Y_0^n = X_0^n = Z_0^n$ for $n \in \{1, \ldots, N_0^K\}$ and $\mu_0^K = \nu_0^K = \frac{1}{K}\sum_{n=1}^{N_0^K}\delta_{Z_0^n}$. We also set two counters: $\overline{N}_0^K = N_0^K$ and $m = 0$, and we define $T_0 = 0$. Last, we initialize $j_0(n) = n$ for all $n \geq 1$.
1. For $t \geq T_m$, we set $j_t(n) = j_{T_m}(n)$ and $\mathrm{d}B_t^n = \mathrm{d}W_t^{j_t(n)}$, $n \geq 1$, and we define the dynamics of the two populations by:

$$X_t^n = X_{T_m} + \int_{T_m}^t b\big(X_s^n, H*\mu_s^K(X_s^n)\big)\,\mathrm{d}s + \int_{T_m}^t \sigma\big(X_s^n, G*\mu_s^K(X_s^n)\big)\mathrm{d}B_s^n,\ n = 1, \ldots, N_{T_m}^K,$$

   and

$$Y_t^n = Y_{T_m} + \int_{T_m}^t b\big(Y_s^n, H*\mu_s(Y_s^n)\big)\,\mathrm{d}s + \int_{T_m}^t \sigma\big(Y_s^n, G*\mu_s(Y_s^n)\big)\mathrm{d}B_s^n, \quad n = 1, \ldots, N_{T_m}^K,$$

   until the first time $t > T_m$ with (t, ρ, θ) an atom of $\mathcal{N}$, such that

$$\rho \leq N_{T_m}^K \quad \text{and} \quad \theta \leq r + c\frac{N_{T_m}^K}{K}.$$

   We then set $T_{m+1} = t$.
2. For $(t, \rho, \theta) = (T_{m+1}, \rho, \theta)$ as before,
   - If $\theta \leq r$, we update $N_t^K \coloneqq N_{t-}^K + 1$ and $\overline{N}_t^K \coloneqq \overline{N}_{t-}^K + 1$, then we define:

$$X_t^{N_t^K} \coloneqq X_{t-}^{\mathbf{i}(\rho)} \quad \text{and} \quad Y_t^{N_t^K} \coloneqq \Lambda_t^{N_{t-}^K}\Big((X_{t-}^n)_{n=1}^{N_{t-}^K}, \rho\Big).$$

   - If $r < \theta \leq r + cN_{T_m}^K/K$, we update $N_t^K \coloneqq N_{t-}^K - 1$, then we redefine:

$$\big(X_t^{\mathbf{i}(\rho)}, X_t^{\mathbf{i}(\rho)+1}, \ldots, X_t^{N_t^K}\big) \coloneqq \big(X_{t-}^{\mathbf{i}(\rho)+1}, X_{t-}^{\mathbf{i}(\rho)+2}, \ldots, X_{t-}^{N_{t-}^K}\big),$$
$$\big(Y_t^{\mathbf{i}(\rho)}, Y_t^{\mathbf{i}(\rho)+1}, \ldots, Y_t^{N_t^K}\big) \coloneqq \big(Y_{t-}^{\mathbf{i}(\rho)+1}, Y_{t-}^{\mathbf{i}(\rho)+2}, \ldots, Y_{t-}^{N_{t-}^K}\big),$$

     and we set $j_t(n) \coloneqq j_{t-}(n+1)$ for all $n \geq \mathbf{i}(\rho)$.
3. We increase m by one and go to Step 1.

Let us explain in words how the algorithm works. The systems $(\mu_t^K)_{t\geq 0}$ and $(\nu_t^K)_{t\geq 0}$ start at time $t=0$ from the same empirical measure, and pairs of particles are given birth or die in the two systems simultaneously from then on. The variable N_t^K counts the current number of living particles in each system at time t. The variable $\overline{N}_t^K$ in turn counts how many particles have been alive in each of the two systems or, equivalently, how many Brownian motions from $(W^j)_{j\geq 1}$ have been used, during the whole time interval $[0,t]$. The usefulness of this counter will come clear shortly.

Now, given an atom (t,ρ,θ), its coordinate t is used to sample a proposal of a birth or dead time, and θ an "action" among those two, according to whether $\theta \leq r$ or $r < \theta \leq r + cN_{t-}^K/K$ respectively.

In a birth event, $\rho \leq N_{t-}^K$ samples two positions in space, one distributed according to $\bar{\mu}_{t-}^K$ for the system μ^K, where $\bar{\mu}_{t-}^K$ is the normalization (see (2.2)) of μ_{t-}^K, and one according to $\bar{\mu}_t$ for the system ν^K, which are optimally coupled as explained before. The pair of newborn particles picks upon birth at time t a new, common driving Brownian motion $(W_s^{\overline{N}^K})_{s\geq t}$ that is independent of the past of the systems.

In a death event, $\rho \leq N_{t-}^K$ samples a uniformly distributed atom from $\bar{\mu}_{t-}^K$ for the system μ^K and from $\bar{\nu}_{t-}^K$ for the system ν^K, with equal index $\mathbf{i}(\rho)$, where $\bar{\nu}_{t-}^K$ is the normalization (see (2.2)) of ν_{t-}^K. The two corresponding particles are then removed, and their common driving Brownian motion, which corresponds to some W^j with $j \leq \overline{N}_t^K$, is discarded forever. The indexes of the particles in the two systems are then updated, as well as the Brownian motions from $(W^j)_{j\geq 1}$ labelled $B^{\mathbf{i}(\rho)}, B^{\mathbf{i}(\rho)+1}, ...$, in order that the particles still alive remain indexed by a full discrete interval of the form $\{1,\dots,N_t^K\}$, and that the underlying Brownian motion W^j driving each pair is preserved. Notice that, due to this updating rule, for all times $t\geq 0$ we have $j_t(N_t^K) = \overline{N}_t^K$.

The system $(\nu_t^K)_{t\geq 0}$ satisfies condition (C.1) by construction. In the next paragraph, we will check that it also satisfies condition (C.2).

3.2 Verification of condition (C.2)

We will denote by $(\mathcal{F}_t)_{t\geq 0}$ the complete filtration generated by all the random objects effectively employed in the algorithm until each time:

$$\mathcal{F}_t \coloneqq \overline{\sigma\left(N_0^K, (Z_0^n)_{n\in\{1,\dots,N_0^K\}}, (\mathcal{N}((0,s],\cdot,\cdot) : s\leq t), (B_s^n : s \leq t)_{n\in\{1,\dots,N_t^K\}}\right)}, \tag{3.4}$$

and by $(\mathcal{G}_t)_{t\geq 0}$ its subfiltration

$$\mathcal{G}_t \coloneqq \overline{\sigma\left(N_s^K : s\leq t\right)}. \tag{3.5}$$

Notice that $\mathcal{N}$ is an $(\mathcal{F}_t)_{t\geq 0}$-Poisson process, and that $(N_t^K)_{t\geq 0}, (\overline{N}_t^K)_{t\geq 0}$ and $(j_t(n) : t\geq 0)$, $n\geq 1$ are processes adapted to $(\mathcal{G}_t)_{t\geq 0}$.

Remark 14. Thanks to Lemma 13, the mapping

$$(t,\omega,\rho) \mapsto \left(\Lambda_t^{N_{t-}^K}\left((X_{t-}^n)_{n=1}^{N_{t-}^K}, \rho\right), X_{t-}^{\mathbf{i}(\rho)}\right) = \left(\Lambda\left(N_{t-}^K, t, (X_{t-}^n)_{n=1}^{N_{t-}^K}, \rho\wedge N_{t-}^K\right), X_{t-}^{\mathbf{i}(\rho)}\right),$$

is measurable with respect to $\mathcal{P}red(\mathcal{F}_t)\otimes\mathcal{B}(\mathbb{R})$, with $\mathcal{P}red(\mathcal{F}_t)\subseteq \mathcal{B}(\mathbb{R})\otimes\mathcal{F}$ the predictable sigma-field associated with $(\mathcal{F}_t)_{t\geq 0}$.

The following identity in law is crucial to check (C.2).

Lemma 15. Assume (H). Let $(\mu_t)_{t\geq 0}$ be the unique weak solution in $\mathcal{M}^+(\mathbb{R}^d)$ of equation (1.2) in Theorem 2, and Λ denote the function constructed in Lemma 13 with the flow of probability measures $(\bar{\mu}_t)_{t\geq 0}$, the normalized version of $(\mu_t)_{t\geq 0}$. Let $(\overline{T}_j)_{j\geq 1}$ denote the sequence of consecutive birth times in $(0,\infty)$ of one new particle in the system $(\nu_t^K)_{t\geq 0}$

constructed with algorithm (A), and $(\overline{T}_j, \rho_j)$ be the first two coordinates of the atom (t,ρ,θ) corresponding to $t = \overline{T}_j$. Then, conditionally on $\mathcal{F}_{\overline{T}_j-}$ and $\left\{\rho_j \le N^K_{\overline{T}_j-}\right\}$,

$$Y^{N^K_{\overline{T}_j}}_{\overline{T}_j} = \Lambda^{N^K_{\overline{T}_j-}}\left((X^n_{t-})_{n=1}^{N^K_{t-}}, \rho_j\right) \text{ has law } \bar{\mu}_{\overline{T}_j}.$$

Proof. Let $f\colon \mathbb{R}^d \to \mathbb{R}$ be a bounded measurable function and $(U_t)_{t\ge 0}$ a bounded $(\mathcal{F}_t)_{t\ge0}$-predictable process (with $\mathcal{F}_t$ given in (3.4)). We have

$$\begin{aligned}
& f\left(Y^{N^K_{\overline{T}_j}}_{\overline{T}_j}\right)\mathbf{1}_{\left\{\rho_j \le N^K_{\overline{T}_j-}\right\}} U_{\overline{T}_j} \\
&\quad = \int_0^\infty\int_0^\infty\int_0^\infty f\left(\Lambda_t^{N^K_{t-}}\left((X^n_{t-})_{n=1}^{N^K_{t-}},\rho\right)\right)\mathbf{1}_{\{\rho\le N^K_{t-},\, \overline{N}^K_{t-}=N^K_0+j-1,\, \theta\le r\}} U_t\, \mathcal{N}(\mathrm{d}t,\mathrm{d}\rho,\mathrm{d}\theta).
\end{aligned}$$

By Remark 14, we can use the compensation formula with respect to the filtration $(\mathcal{F}_t)_{t\ge0}$, and deduce with Lemma 13 that

$$\begin{aligned}
&\mathbb{E}\left(f\left(Y^{N^K_{\overline{T}_j}}_{\overline{T}_j}\right)\mathbf{1}_{\left\{\rho_j \le N^K_{\overline{T}_j-}\right\}} U_{\overline{T}_j}\right) \\
&\qquad = \int_0^\infty\int_0^\infty \mathbb{E}\left(\langle\bar{\mu}_t, f\rangle N^K_t \mathbf{1}_{\{\overline{N}^K_t = N^K_0+j-1,\, \theta\le r\}} U_t\right)\mathrm{d}\theta\mathrm{d}t \\
&\qquad = \mathbb{E}\left(\int_{[0,\infty)^3}\langle\bar{\mu}_t, f\rangle\mathbf{1}_{\{\rho\le N^K_{t-},\, \overline{N}^K_{t-}=N^K_0+j-1,\, \theta\le r\}} U_t\, \mathcal{N}(\mathrm{d}t,\mathrm{d}\rho,\mathrm{d}\theta)\right) \\
&\qquad = \mathbb{E}\left(\langle\bar{\mu}_{\overline{T}_j}, f\rangle\mathbf{1}_{\left\{\rho_j\le N^K_{\overline{T}_j-}\right\}} U_{\overline{T}_j}\right).
\end{aligned}$$

Since any bounded random variable measurable w.r.t. $\mathcal{F}_{\overline{T}_j-}$ can be written as $U_{\overline{T}_j}$ for some predictable process $(U_t)_{t\ge0}$, the statement is proved. ☐

Proposition 16. Assume (H). For each $t \ge 0$, conditionally on $\langle\nu^K_t, 1\rangle$, the $K\langle\nu^K_t,1\rangle$ atoms of the measure ν^K_t constructed in algorithm (A) are i.i.d. random variables of law $\bar{\mu}_t = \mu_t/\langle\mu_t,1\rangle$, with $(\mu_s)_{s\ge0}$ the unique weak solution in $\mathcal{M}^+(\mathbb{R}^d)$ of equation (1.2) given by Theorem 2. That is to say, condition (C.2) holds.

Proof. The proof will be done constructing an alternative system $(\widehat{\nu}^K_t = \frac{1}{K}\sum_{n=1}^{N^K_t}\delta_{\widehat{Y}^n_t})_{t\ge0}$ with the same law as $(\nu^K_t)_{t\ge0}$, for which the required property is easily checked. This system is defined on the same probability space as $(\nu^K_t)_{t\ge0}$, by means of a variant of the construction of $(\nu^K_t)_{t\ge0}$ in algorithm (A). This algorithm is as follows:

0. Define for all $j\ge1$:

$$Z^j_t = Z^j_0 + \int_0^t b\big(Z^j_s, H*\mu_s(Z^j_s)\big)\,\mathrm{d}s + \int_0^t \sigma\big(Z^j_s, G*\mu_s(Z^j_s)\big)\,\mathrm{d}W^j_s\,, \quad t\ge0.$$

   Set $\widehat{Y}^n_0 = Z^n_0$ for $n\in\{1,\dots,N^K_0\}$ and $\widehat{\nu}^K_0 = \frac{1}{K}\sum_{n=1}^{N^K_0}\delta_{\widehat{Y}^n_0}$. As before, we set the same counters $\overline{N}^K_0 = N^K_0$ and $m=0$, we define $T_0=0$ and we initialize $j_0(n)=n$ for all $n\ge1$.
1. For $t\ge T_m$, we set $j_t(n) = j_{T_m}(n)$ and $\mathrm{d}B^n_t = \mathrm{d}W^{j_t(n)}_t$, $n\ge1$, and we take

$$\widehat{Y}^n_t = Z^{j_t(n)}_t, \quad n=1,\dots,N^K_{T_m},$$

   until the first time $t > T_m$ with (t,ρ,θ) an atom of $\mathcal{N}$, such that $\rho\le N^K_{T_m}$ and $\theta\le r + cN^K_{T_m}/K$. We then set $T_{m+1}=t$.

2. For $(t,\rho,\theta)=(T_{m+1},\rho,\theta)$ as before,
   - If $\theta \le r$, we update $N_t^K := N_{t-}^K + 1$ and $\overline{N}_t^K := \overline{N}_{t-}^K + 1$, then we define:
   $$\widehat{Y}_t^{N_t^K} := Z_t^{\overline{N}_t^K}.$$
   - If $r < \theta \le r + cN_{T_m}^K/K$, we update $N_t^K := N_{t-}^K - 1$, and we redefine:
   $$\left(\widehat{Y}_t^{\mathbf{i}(\rho)}, \widehat{Y}_t^{\mathbf{i}(\rho)+1}, \dots, \widehat{Y}_t^{N_t^K}\right) := \left(\widehat{Y}_{t-}^{\mathbf{i}(\rho)+1}, \widehat{Y}_{t-}^{\mathbf{i}(\rho)+2}, \dots, \widehat{Y}_{t-}^{N_{t-}^K}\right),$$
   and $j_t(n) := j_{t-}(n+1)$ for all $n \ge \mathbf{i}(\rho)$.
3. We increase m by one and go to Step 1.

Plainly, instead of sampling at each birth time $\overline{T}_j$ the position of a new independent particle $Y^{N_{\overline{T}_j}^K}$ from the atom $(\overline{T}_j,\rho,\theta)$ of $\mathcal{N}$ as in (A), we now add a new particle $\widehat{Y}^{N_{\overline{T}_j}^K}$ to the system by "turning on" at that time the nonlinear diffusion process $Z^{\overline{N}_{\overline{T}_j}^K} = Z^{N_0^K+j}$, which has evolved independently since time $t=0$, driven by the same Brownian motion $W^{N_0^K+j}$ that drives the process $\left(Y_t^{N_{\overline{T}_j}^K} : t \ge \overline{T}_j\right)$ in the construction (A). Call now

$$\widehat{\mathcal{F}}_t := \overline{\sigma\left(\mathcal{F}_t \vee \left(Z_{\overline{T}_k}^{N_0^K+k} : N_0^K + k \le \overline{N}_t^K\right)\right)},$$

(with $\mathcal{F}_t$ as in (3.4)) the filtration containing the information effectively employed to construct the process $(\widehat{\nu}_t^K)$, and let $(V_t)_{t\ge 0}$ be a bounded left continuous process adapted to $(\widehat{\mathcal{F}}_t)_{t\ge0}$. Conditionally on N_0^K, $V_{\overline{T}_j}$ depends only on $\mathcal{N}$ and (W^k, Z_0^k) for $k < N_0^K + j$, while $\left(Z_t^{N_0^K+j}\right)_{t\ge0}$ is independent of them. Therefore, we have

$$\begin{aligned}\mathbb{E}\left(f\left(\widehat{Y}_{\overline{T}_j}^{N_{\overline{T}_j}^K}\right)\mathbf{1}_{\left\{\rho_j \le N_{\overline{T}_j-}^K\right\}} V_{\overline{T}_j}\right) &= \mathbb{E}\left(f\left(Z_{\overline{T}_j}^{N_0^K+j}\right)\mathbf{1}_{\left\{\rho_j \le N_{\overline{T}_j-}^K\right\}} V_{\overline{T}_j}\right)\\ &= \mathbb{E}\left(\langle \bar{\mu}_{\overline{T}_j}, f\rangle \mathbf{1}_{\left\{\rho_j \le N_{\overline{T}_j-}^K\right\}} V_{\overline{T}_j}\right),\end{aligned}$$

by Remark 12 a). This implies that, conditionally on $\widehat{\mathcal{F}}_{\overline{T}_j-}$ and $\{\rho_j \le N_{\overline{T}_j-}^K\}$, the random variable $\widehat{Y}_{\overline{T}_j}^{N_{\overline{T}_j}^K}$ has the law $\bar{\mu}_{\overline{T}_j}$. Comparing this to the setting in Lemma 15, one can check by induction on j that the processes $(\nu_t^K)_{t\ge0}$ and $(\widehat{\nu}_t^K)$ have the same law on each of their (common) time intervals $[0,\overline{T}_j]$, hence over all $[0,\infty)$.

To conclude, notice that the i.i.d processes $(Z_t^j)_{t\ge0}, j\ge1$ have law $\bar{\mu}_t$ at each $t\ge0$, and they are independent of the filtration $(\mathcal{G}_t)_{t\ge0}$ defined in (3.5), with respect to which the process $(N_t^K)_{t\ge0}$ is measurable. Moreover, for each $t\ge0$, $\{\widehat{Y}_t^1,\dots,\widehat{Y}_t^{N_t^K}\} = \{Z_t^{j_t(1)},\dots,Z_t^{j_t(N_t^K)}\}$ is a random subset of $\{Z_t^1,\dots,Z_t^{\overline{N}_t^K}\}$, selected in a way that is measurable w.r.t. $\mathcal{G}_t$. This readily implies that, conditionally on $\{N_t^K = N\}$, $\{\widehat{Y}_t^1,\dots,\widehat{Y}_t^N\}$ are N i.i.d. random variables of law $\bar{\mu}_t$, as required. □

4 Proof of Theorem 3: pure binary branching case

We consider in this section the case where interactions take place only through the reproduction events, that is, due only to the fact that the position of a newborn individual coincides at its birth with that of its parent (after which all individuals evolve completely independently). We provide the complete proof for this case as it might be of independent interest, since convergence bounds are neither available in this basic setting, and also because it is useful to illustrate directly the main arguments.

We assume the following throughout this section.

Hypothesis (H'):

H.1'. (H.1) holds.

H.2'. The coefficients $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\otimes d}$ and $b\colon \mathbb{R}^d \to \mathbb{R}^d$ do not depend on μ_t^K and, moreover, they are Lipschitz continuous with σ bounded (for simplicity).

H.3'. The individual instantaneous birth and death rates are time inhomogeneous, specified by two measurable functions $r, c\colon [0,T] \to \mathbb{R}_+$ bounded by some positive constants $\bar{r}$ and $\bar{c}$, respectively.

Notice that, since r and c are deterministic measurable functions of t, they are predictable when seen as processes (cf. the sigma-field generated by continuous functions on $\mathbb{R}_+$ is the Borel sigma-field).

The analog of Theorem 2 is standard in this scenario (or can be proved by the same techniques used in [12]), and the limit in law of the process $(\mu_t^K)_{t\geq 0}$ is given by the unique weak solution in $\mathcal{M}^+(\mathbb{R}^d)$ to the linear evolution equation with initial condition μ_0:

$$\langle \mu_t, f(t,\cdot)\rangle = \langle \mu_0, f(0,\cdot)\rangle + \int_0^t \langle \mu_s, \partial_s f(s,\cdot) + Lf(s,\cdot) + (r(s)-c(s))f(s,\cdot)\rangle \,\mathrm{d}s, \quad \forall t\in[0,T], \tag{4.1}$$

for each $f \in C^{1,2}([0,T]\times\mathbb{R}^d)$, where L is the time-homogeneous operator on $C^2(\mathbb{R}^d)$ functions ϕ:

$$L\phi(x) = \frac{1}{2}\mathrm{Tr}\big(a(x)\mathrm{Hess}\phi(x)\big) + b(x)\cdot\nabla\phi(x). \tag{4.2}$$

The construction of the coupling with the auxiliary system is essentially the same as in Section 3, using algorithm (A) with two minor modifications:

- Step 1 is carried out until the first time $t > T_m$, where (t,ρ,θ) is an atom of $\mathcal{N}$ such that $\rho \leq N_{T_m}^K$ and $\theta \leq r(t)+c(t)$, at which time we set $T_{m+1} = t$.
- The updates in Step 2 are carried out according to whether $\theta \leq r(t)$ or otherwise $r(t) < \theta \leq r(t)+c(t)$.

In between birth or deaths events, individuals X^n in the system $(\mu_t^K)_{t\geq 0}$ evolve according to the SDEs

$$\mathrm{d}X_t^n = b(X_t^n)\,\mathrm{d}t + \sigma(X_t^n)\,\mathrm{d}B_t^n, \quad n = 1,\dots,N_t^K,$$

as also do the individuals Y^n in the system $(\nu_t^K)_{t\geq 0}$.

We establish some controls for the mass of the process $(\mu_t^K)_{t\geq 0}$.

Lemma 17. Under assumption (H'), for each $T>0$ and $p\geq 1$ there is a constant $C_{T,p}>0$ such that

$$\sup_{K\in\mathbb{N}\setminus\{0\}} \mathbb{E}\left(\sup_{t\in[0,T]} \langle \mu_t^K, 1\rangle^p\right) < C_{T,p} \sup_{K\in\mathbb{N}\setminus\{0\}} \mathbb{E}(\langle \mu_0^K, 1\rangle^p).$$

Moreover, if $\sup_{K\in\mathbb{N}\setminus\{0\}} \mathbb{E}(\langle \mu_0^K, 1\rangle) < \infty$, for all $T>0$ we have

$$\mathbb{E}\big(\big|\langle \mu_t^K, 1\rangle - \langle \mu_t, 1\rangle\big|\big) \leq C_T\big(I_1(K) + K^{-\frac{1}{2}}\big), \quad \forall t\in[0,T],$$

for some $C_T>0$, with

$$I_1(K) = \mathbb{E}\big(|\langle \mu_0^K, 1\rangle - \langle \mu_0, 1\rangle|\big).$$

Proof. The first claim is shown as in [12, Lemma 3.3] in a more general setting. For the second assertion, we write the dynamics of the number of particles in the system in terms of the Poisson point measure $\mathcal{N}$ used in algorithm (A). We obtain for all $t\geq 0$ that

$$N_t^K = N_0^K + \int_0^t \int_{\mathbb{R}_+}\int_{\mathbb{R}_+} \mathbf{1}_{\rho\leq N_{s-}^K}\left(\mathbf{1}_{\theta\leq r(s)} - \mathbf{1}_{r(s)<\theta\leq r(s)+c(s)}\right)\mathcal{N}(\mathrm{d}s,\mathrm{d}\rho,\mathrm{d}\theta)$$

$$= N_0^K + \int_0^t (r(s)-c(s))N_s^K \,\mathrm{d}s + M_t^K,$$

where $(M_t^K)_{t\geq 0}$ is a martingale since, for all $t \geq 0$,

$$\mathbb{E}\Big(\int_0^t \int_{\mathbb{R}_+}\int_{\mathbb{R}_+} \Big|\mathbf{1}_{\rho\leq N_s^K}\Big(\mathbf{1}_{\theta\leq r(s)} - \mathbf{1}_{r(s)<\theta\leq r(s)+c(s)}\Big)\Big| \,\mathrm{d}s\mathrm{d}\rho\mathrm{d}\theta\Big) \leq (\bar{r}+\bar{c})\mathbb{E}\Big(\int_0^t N_s^K \,\mathrm{d}s\Big) < \infty,$$

by the first part and the assumption on the total mass. Comparing this evolution to the ODE (4.3) satisfied by the total mass of the limiting measure, we get the estimate

$$\mathbb{E}\Big(\Big|\frac{N_t^K}{K} - \langle \mu_t, 1\rangle\Big|\Big) \leq \mathbb{E}\Big(\Big|\frac{N_0^K}{K} - \langle \mu_0, 1\rangle\Big|\Big) + (\bar{r}+\bar{c})\int_0^t \mathbb{E}\Big(\Big|\frac{N_s^K}{K} - \langle \mu_s, 1\rangle\Big|\Big)\,\mathrm{d}s + \mathbb{E}\Big(\frac{|M_t^K|}{K}\Big).$$

The last term is controlled using the Burkholder-Davis-Gundy (BDG) inequality as follows

$$\begin{aligned}
\mathbb{E}\Big(\frac{|M_t^K|}{K}\Big) &\leq \frac{1}{K}\,\mathbb{E}\left(\int_0^t \int_{\mathbb{R}_+}\int_{\mathbb{R}_+} \mathbf{1}_{\{\rho\leq N_{s-}^K,\, \theta\leq r(s)+c(s)\}}\,\mathcal{N}(\mathrm{d}s,\mathrm{d}\rho,\mathrm{d}\theta)\right)^{\frac{1}{2}} \\
&= \frac{\mathbb{E}((\int_0^t (r(s)+c(s))N_s^K \,\mathrm{d}s)^{\frac{1}{2}}}{K} \\
&\leq \frac{C_T}{\sqrt{K}}\left(\sup_{K\in\mathbb{N}\setminus\{0\}} \mathbb{E}(\langle \mu_0^K, 1\rangle)(\bar{r}+\bar{c})e^{\bar{r}}t\right)^{\frac{1}{2}},
\end{aligned}$$

for all $t \in [0,T]$. We conclude by Gronwall's lemma that

$$\mathbb{E}\Big(\Big|\frac{N_t^K}{K} - \langle \mu_t, 1\rangle\Big|\Big) \leq C_T\Big(\mathbb{E}\Big(\Big|\frac{N_0^K}{K} - \langle \mu_0, 1\rangle\Big|\Big) + \frac{1}{\sqrt{K}}\Big), \quad \forall t \in [0,T]. \quad \square$$

The analogue of Proposition 11 in this section's setting is rather elementary, yet illustrative for the general case, so we state it in full details and prove it next.

Proposition 18. Assume (H') and let $(\mu_t)_{t\geq 0}$ be the unique weak solution in $\mathcal{M}^+(\mathbb{R}^d)$ of the linear equation (4.1) with initial condition μ_0, written for short as

$$\frac{\partial \mu_t}{\partial t} = L^*\mu_t + (r(t)-c(t))\mu_t, \quad t \geq 0,$$

with L^* the adjoint of the operator L given in (4.2). Let $(Y_t)_{t\geq 0}$ be the unique pathwise solution to the SDE

$$Y_t = Y_0 + \int_0^t b(Y_s)\,\mathrm{d}s + \int_0^t \sigma(Y_s)\,\mathrm{d}W_s, \quad t \geq 0,$$

where W is a d-dimensional Brownian motion and Y_0 an independent random variable in $\mathbb{R}^d$ with law $\tilde{\mu}_0 = \bar{\mu}_0$. Then, the flow $(\tilde{\mu}_t)_{t\geq 0}$ of time-marginal laws of $(Y_t)_{t\geq 0}$ is the unique weak solution of the Fokker-Planck equation

$$\frac{\partial \tilde{\mu}_t}{\partial t} = L^*\tilde{\mu}_t, \quad t \geq 0,$$

with initial condition $\tilde{\mu}_0 = \bar{\mu}_0$ and one has $\tilde{\mu}_t = \bar{\mu}_t$ for all $t \geq 0$. In particular, for each bounded real function f we have $\langle \mu_t, f\rangle = \mathbb{E}(f(Y_t)n_t)$, where n_t is the unique solution with $n_0 = \langle \mu_0, 1\rangle$ of the linear differential equation

$$\mathrm{d}n_t = (r(t)-c(t))n_t\,\mathrm{d}t. \tag{4.3}$$

Proof. The first claim is standard and easily seen using Itô's formula (uniqueness is also standard using e.g. the Feynman-Kac formula). The relation between the law of Y_t and μ_t for all $t \geq 0$ is easily shown considering the function $h(t,x) = \langle \mu_t, 1\rangle f(t,x)$ and computing

$$\begin{aligned}
&\langle \tilde{\mu}_t, h(t,\cdot)\rangle \\
&\qquad = \langle \tilde{\mu}_0, h(0,\cdot)\rangle + \int_0^t \langle \tilde{\mu}_s, \partial_s h(s,\cdot) + Lh(s,\cdot)\rangle \,\mathrm{d}s \\
&\qquad = \langle \langle \mu_0, 1\rangle \tilde{\mu}_0, f(0,\cdot)\rangle + \int_0^t \langle \tilde{\mu}_s, f(s,\cdot)\partial_s\langle \mu_s, 1\rangle + \langle \mu_s, 1\rangle \partial_s f(s,\cdot) + \langle \mu_s, 1\rangle Lf(s,\cdot)\rangle \,\mathrm{d}s \\
&\qquad = \langle \langle \mu_0, 1\rangle \tilde{\mu}_0, f(0,\cdot)\rangle + \int_0^t \langle \langle \mu_s, 1\rangle \tilde{\mu}_s, \partial_s f(s,\cdot) + Lf(s,\cdot) + (r(s) - c(s)) f(s,\cdot)\rangle \,\mathrm{d}s.
\end{aligned}$$

This means that $(\langle \mu_t, 1\rangle \tilde{\mu}_t)_{t\geq 0}$ satisfies equation (4.1). Uniqueness for that equation yields $\langle \mu_t, 1\rangle \tilde{\mu}_t = \mu_t$ for all $t \geq 0$ as claimed. Consequently,

$$\langle \mu_t, f\rangle = \mathbb{E}(\langle \mu_t, 1\rangle f(Y_t)),$$

for all bounded f, and the fact that $(\langle \mu_t, 1\rangle)_{t\geq 0}$ satisfies (4.3) is immediate. $\square$

In order to prove that condition (C.3) holds, one last additional estimate is needed, which will be used to control the joint evolution of coupled particles, in between birth or death events.

Lemma 19. Assume (H.2') and let $X = (X_t)_{t\geq 0}$ and $Y = (Y_t)_{t\geq 0}$ be two diffusion processes with generator L given in (4.2), both driven by a given Brownian motion B in $\mathbb{R}^d$. For each $T > 0$ there exists $C_T > 0$ such that for all $0 < u < t < T$

$$\mathbb{E}(\|X_t - Y_t\|^2 - \|X_u - Y_u\|^2) \leq C_T \int_u^t \mathbb{E}(\|X_s - Y_s\|^2) \,\mathrm{d}s.$$

Proof. Let $(\tau_n)_{n\in\mathbb{N}}$ be the sequence defined by $\tau_n \coloneqq \inf\{s \geq 0 : \|X_s\|^2 + \|Y_s\|^2 > n\}$, which localizes the local martingale parts of X and Y. We first establish a control on the running suprema of the processes. Using the fact that b is Lipschitz we obtain

$$\begin{aligned}
\sup_{u\in[0,t\wedge\tau_n]} \|X_u\|^2 &\leq 2\|X_0\|^2 + C_T + C_T \int_0^t \sup_{u\in[0,s\wedge\tau_n]} \|X_u\|^2 \,\mathrm{d}s \\
&\qquad + 2\sum_{i,j=1}^d \left(\sup_{u\in[0,t\wedge\tau_n]} \left| \int_0^u \sigma^{(ij)}(X_s) \,\mathrm{d}B_s^{(j)} \right| \right)^2.
\end{aligned}$$

With the BDG inequality and the fact that σ is also Lipschitz we then get

$$\mathbb{E}\left(\sup_{u\in[0,t\wedge\tau_n]} \|X_u\|^2 \right) \leq 2\mathbb{E}\left(\|X_0\|^2\right) + C_T + C_T \int_0^t \mathbb{E}\left(\sup_{u\in[0,s\wedge\tau_n]} \|X_u\|^2 \right) \mathrm{d}s.$$

Applying Gronwall's lemma and then Fatou's lemma upon letting $n \to \infty$ we deduce

$$\mathbb{E}\left(\sup_{s\in[0,T]} \|X_s\|^2 \right) \leq C_T(\mathbb{E}(\|X_0\|^2) + 1), \tag{4.4}$$

and a similar estimate holds for the process Y. Now, Itô's formula shows that

$$\|X_t - Y_t\|^2 = \|X_u - Y_u\|^2 + \int_u^t 2(X_s - Y_s)^{\mathrm{t}}(b(X_s) - b(Y_s)) \,\mathrm{d}s$$

$$+ \int_u^t 2(X_s - Y_s)^{\mathrm{t}}(\sigma(X_s) - \sigma(Y_s))\,\mathrm{d}B_s + \sum_{i,j=1}^d \int_u^t (\sigma^{(ij)}(X_s) - \sigma^{(ij)}(Y_s))^2\,\mathrm{d}s.$$

The sequence $(\tau_n)_n$ localizes the local martingale on the right hand side. Taking expectation for the stopped process and using the Lipschitz character of b and σ leads to

$$\mathbb{E}(\|X_{t\wedge\tau_n} - Y_{t\wedge\tau_n}\|^2) \le \mathbb{E}(\|X_u - Y_u\|^2) + C\int_u^t \mathbb{E}(\|X_{s\wedge\tau_n} - Y_{s\wedge\tau_n}\|^2)\,\mathrm{d}s.$$

By dominated convergence using the bound (4.4), we can take $n \to \infty$ and conclude. □

Now we can state the bound leading to condition (C.3), and to the proof of the main result, in the case of pure binary branching.

Lemma 20. Assume (H') and let $(\bar{\mu}_t)_{t\ge 0}$ and $(\bar{\nu}_t^K)_{t\ge 0}$ be the normalizations (see (2.2)) of the solution $(\mu_t)_{t\ge 0}$ of (4.1) and of the process $(\nu_t^K)_{t\ge 0}$ constructed using algorithm (A) (modified as mentioned at the beginning of the section) respectively. Then, there exists a constant $C_T > 0$ depending on d and q, such that for all $K \in \mathbb{N}\setminus\{0\}$ and $t \in [0,T]$

$$\mathbb{E}\left(\frac{1}{K}\sum_{n=1}^{N_t^K} \|X_t^n - Y_t^n\|^2\right) \le C_T \int_0^t \mathbb{E}\left(\frac{N_s^K}{K} W_2^2(\bar{\nu}_s^K, \bar{\mu}_s)\right)\mathrm{d}s.$$

Proof. Consider the product empirical measure $\eta_t^K \coloneqq \frac{1}{K}\sum_{n=1}^{N_t^K} \delta_{(X_t^n, Y_t^n)}$ and the sequence of jump times $(T_m)_{m\in\mathbb{N}}$ of the process $(N_t^K)_{t\ge 0}$, defined through algorithm (A). We decompose the evolution of η_t^K in terms of $(T_m)_{m\in\mathbb{N}}$ as follows

$$\eta_t^K = \eta_0^K + \eta_t^K - \eta_{T_{A_t^K}}^K + \sum_{m=1}^\infty \mathbf{1}_{t\ge T_m}\left(\eta_{T_m}^K - \eta_{T_m^-}^K + \eta_{T_m^-}^K - \eta_{T_{m-1}}^K\right),$$

where $A_t^K \coloneqq \sum_{s\le t} |\Delta N_s^K|$ with $\Delta N_s^K = N_s^K - N_{s-}^K$. The aim of this decomposition is to control separately what happens in between jumps and at the jump instants. Integrating the function $d_2(x,y) \coloneqq \|x-y\|^2$ and taking expectation yields

$$\begin{aligned}\mathbb{E}\big(\langle \eta_t^K, d_2\rangle\big) &= \mathbb{E}\big(\langle \eta_0^K, d_2\rangle\big) + \mathbb{E}\left(\sum_{m=1}^\infty \mathbf{1}_{t\ge T_m}\Big(\langle \eta_{T_m}^K, d_2\rangle - \langle \eta_{T_m^-}^K, d_2\rangle\Big)\right)\\ &\quad + \mathbb{E}\left(\langle \eta_t^K, d_2\rangle - \left\langle \eta_{T_{A_t^K}}^K, d_2\right\rangle + \sum_{m=1}^\infty \mathbf{1}_{t\ge T_m}\left(\langle \eta_{T_m^-}^K, d_2\rangle - \langle \eta_{T_{m-1}}^K, d_2\rangle\right)\right).\end{aligned} \tag{4.5}$$

By Lemma 19, and since the evolution of η_t^K is independent of the sigma-field $(\mathcal{G}_t)_{t\ge 0}$ (defined in (3.5)) on each interval $[T_{m-1}, T_m)$, we get

$$\mathbb{E}\big(\mathbf{1}_{t\ge T_m}(\langle \eta_{T_m^-}^K, d_2\rangle - \langle \eta_{T_{m-1}}^K, d_2\rangle) \mid \mathcal{G}_t\big) \tag{4.6}$$

$$\begin{aligned}&= \mathbb{E}\left(\frac{1}{K}\sum_{n=1}^{N_{T_{m-1}}^K} \|X_{T_m^-}^n - Y_{T_m^-}^n\|^2 - \|X_{T_{m-1}}^n - Y_{T_{m-1}}^n\|^2 \,\middle|\, \mathcal{G}_t\right)\mathbf{1}_{t\ge T_m}\\ &\le \frac{1}{K}\sum_{n=1}^{N_{T_{m-1}}^K} C\int_{T_{m-1}}^{T_m^-} \mathbb{E}\big(\|X_s^n - Y_s^n\|^2 \mid \mathcal{G}_t\big)\,\mathrm{d}s\mathbf{1}_{t\ge T_m}\\ &= C\int_{T_{m-1}}^{T_m^-} \mathbb{E}\big(\langle \eta_s^K, d_2\rangle \mid \mathcal{G}_t\big)\,\mathrm{d}s\mathbf{1}_{t\ge T_m},\end{aligned} \tag{4.7}$$

and similarly, for the remaining time interval,

$$\mathbb{E}\Big(\mathbb{E}\Big(\langle \eta_t^K, d_2\rangle - \langle \eta_{T_{A_t^K}}^K, d_2\rangle \;\Big|\; \mathcal{G}_t\Big)\Big) \leq C\int_{T_{A_t^K}}^t \mathbb{E}\big(\langle \eta_s^K, d_2\rangle \mid \mathcal{G}_t\big)\,\mathrm{d}s.$$

Recalling Step 2 of the variant of algorithm (A) used in this section, the term involving the jumps of the processes can be written as

$$\begin{aligned}
&\mathbb{E}\Big(\sum_{m=1}^{\infty}\mathbf{1}_{t\geq T_m}\,\Big(\langle \eta_{T_n}^K, d_2\rangle - \langle \eta_{T_n^-}^K, d_2\rangle\Big)\Big)\\
&\qquad = \mathbb{E}\Big(\frac{1}{K}\int_{[0,t]\times\mathbb{R}_+\times\mathbb{R}_+}\Big(\mathbf{1}_{\rho\leq N_{s-}^K}\mathbf{1}_{\theta\leq r(s)}\Big\|X_s^{N_s^K}-Y_s^{N_s^K}\Big\|^2\\
&\qquad\qquad - \mathbf{1}_{\rho\leq N_{s-}^K}\mathbf{1}_{r(s)<\theta\leq r(s)+c(s)}\Big\|X_{s-}^{\mathbf{i}(\rho)}-Y_{s-}^{\mathbf{i}(\rho)}\Big\|^2\Big)\,\mathcal{N}(\mathrm{d}s,\mathrm{d}\rho,\mathrm{d}\theta)\Big)\\
&\qquad \leq \mathbb{E}\Big(\int_{[0,t]\times\mathbb{R}_+\times\mathbb{R}_+}\frac{1}{K}\mathbf{1}_{\rho\leq N_{s-}^K}\mathbf{1}_{\theta\leq r(s)}\Big\|X_{s-}^{\mathbf{i}(\rho)}-\Lambda_s^{N_{s-}^K}\Big((X_{s-}^n)_{n=1}^{N_{s-}^K},\rho\Big)\Big\|^2\,\mathcal{N}(\mathrm{d}s,\mathrm{d}\rho,\mathrm{d}\theta)\Big)\\
&\qquad = \mathbb{E}\Big(\int_0^t\frac{N_s^K}{K}\,r(s)W_2^2\big(\bar{\mu}_s^K,\bar{\mu}_s\big)\,\mathrm{d}s\Big),
\end{aligned}\tag{4.8}$$

where we used Lemma 13 and Remark 14 in the last equality. Since $\mathbb{E}(\langle\eta_0^K, d_2\rangle) = 0$, combining the two previous estimates and writing C for some constant that may change from line to line, we deduce

$$\begin{aligned}
\mathbb{E}(\langle\eta_t^K, d_2\rangle) &\leq C\int_0^t \mathbb{E}(\langle\eta_s^K, d_2\rangle)\,\mathrm{d}s + \mathbb{E}\Big(\int_0^t\frac{N_s^K}{K}r(s)W_2^2\big(\bar{\mu}_s^K,\bar{\mu}_s\big)\,\mathrm{d}s\Big)\\
&\leq C\int_0^t \mathbb{E}(\langle\eta_s^K, d_2\rangle)\,\mathrm{d}s + C\int_0^t\mathbb{E}\Big(\frac{N_s^K}{K}W_2^2\big(\bar{\nu}_s^K,\bar{\mu}_s\big)\Big)\,\mathrm{d}s\\
&\qquad + C\int_0^t\mathbb{E}\Big(\frac{N_s^K}{K}W_2^2\big(\bar{\mu}_s^K,\bar{\nu}_s^K\big)\Big)\,\mathrm{d}s\\
&\leq C\int_0^t \mathbb{E}(\langle\eta_s^K, d_2\rangle)\,\mathrm{d}s + C\int_0^t\mathbb{E}\Big(\frac{N_s^K}{K}W_2^2\big(\bar{\nu}_s^K,\bar{\mu}_s\big)\Big)\,\mathrm{d}s,
\end{aligned}$$

where in the last inequality, we used the fact that

$$\mathbb{E}\Big(\frac{N_t^K}{K}W_2^2\big(\bar{\mu}_t^K,\bar{\nu}_t^K\big)\Big) \leq \mathbb{E}\Big(\frac{1}{K}\sum_{n=1}^{N_t^K}\|X_t^n - Y_t^n\|^2\Big),\tag{4.9}$$

since $W_2^2\big(\bar{\mu}_t^K,\bar{\nu}_t^K\big) \leq \frac{1}{N_t^K}\sum_{n=1}^{N_t^K}\|X_t^n-Y_t^n\|^2$. We conclude by Gronwall's lemma. ☐

Corollary 21. Under (H') condition (C.3) holds with the improved bound: $C_T R_{d,q}(K)$.

Proof. Combine inequality (4.9) with Lemma 20 and apply then Lemma 6. ☐

We now have everything that is needed to prove our main result in the case of pure branching diffusions.

Proof of Theorem 3 under (H'). Since $\langle\mu_t^K, 1\rangle = N_t^K/K$, applying Lemma 4 and the triangle inequality for W_1 we get for all $t\in[0,T]$ that

$$\begin{aligned}
\mathbb{E}\left(\|\mu_t^K-\mu_t\|_{\mathrm{BL}^*}\right) &\leq \mathbb{E}\Big(\frac{N_t^K}{K}W_1\big(\bar{\nu}_t^K,\bar{\mu}_t^K\big)\Big) + \mathbb{E}\Big(\frac{N_t^K}{K}W_1\big(\bar{\nu}_t{}^K,\bar{\mu}_t\big)\Big) + \mathbb{E}\big(\big|\langle\mu_t^K,1\rangle - \langle\mu_t,1\rangle\big|\big)\\
&\leq \Big(\mathbb{E}\Big(\frac{N_t^K}{K}W_2^2\big(\bar{\nu}_t^K,\bar{\mu}_t\big)\Big)^{\frac{1}{2}} + \mathbb{E}\Big(\frac{N_t^K}{K}W_2^2\big(\bar{\nu}_t^K,\bar{\mu}_t^K\big)\Big)^{\frac{1}{2}}\Big)\mathbb{E}\Big(\frac{N_t^K}{K}\Big)^{\frac{1}{2}}
\end{aligned}$$

$$+\,\mathbb{E}\big(\big|\langle\mu_t^K,1\rangle-\langle\mu_t,1\rangle\big|\big), \tag{4.10}$$

where we also used the Cauchy-Schwarz inequality and the inequality $W_1^2 \le W_2^2$ in the second line. Thanks to Lemma 17 we obtain, for all $t\in[0,T]$,

$$\mathbb{E}\left(\|\mu_t^K-\mu_t\|_{\mathrm{BL}^*}\right) \le C_T\bigg(\mathbb{E}\Big(\frac{N_t^K}{K}W_2^2\big(\bar\nu_t^K,\bar\mu_t\big)\Big)^{\frac12}+\mathbb{E}\Big(\frac{N_t^K}{K}W_2^2\big(\bar\nu_t^K,\bar\mu_t^K\big)\Big)^{\frac12}+I_1(K)+K^{-1/2}\bigg).$$

Now, thanks to the first bound in Lemma 17, the uniform moment control stated in Remark 12 b), and conditions (C.1) and (C.2), we can apply Lemma 6 to $\bar\nu=\bar\nu_t^K$, $N=N_t^K$, and $\bar\mu=\bar\mu_t$ to bound the first term in the right hand side by $R_{d,q}^{1/2}(K)$. The second term is bounded by $C_T R_{d,q}^{1/2}(K)$, due to Corollary 21. Since $K^{-1/2}\le R_{d,q}^{1/2}$, we conclude that

$$\mathbb{E}\left(\|\mu_t^K-\mu_t\|_{\mathrm{BL}^*}\right) \le C_T\Big(R_{d,q}(K)^{\frac12}+I_1(K)\Big),\quad \forall t\in[0,T]. \qquad \square$$

5 Proof of Theorem 3: general case

We now consider processes $(\mu_t^K)_{t\ge0}$ satisfying the general assumptions of Theorem 3. We start by establishing bounds for the mass of the process, analogous to the bounds in Lemma 17. The convergence bound is less sharp and more difficult to establish now because of the nonlinearities coming from the interaction.

Lemma 22. Assume (H) and let $(\mu_t)_{t\ge0}$ be the unique solution in $\mathcal{M}^+(\mathbb{R}^d)$ of (1.2). For each $T>0$ and $p\ge1$ there is a constant $C_{T,p}>0$ such that

$$\sup_{K\in\mathbb{N}\setminus\{0\}}\mathbb{E}\bigg(\sup_{t\in[0,T]}\langle\mu_t^K,1\rangle^p\bigg) < C_{T,p}\sup_{K\in\mathbb{N}\setminus\{0\}}\mathbb{E}(\langle\mu_0^K,1\rangle^p).$$

Moreover, if $\sup_{K\in\mathbb{N}\setminus\{0\}}\mathbb{E}(\langle\mu_0^K,1\rangle^4)<\infty$, for all $T>0$ we have

$$\mathbb{E}\Big(\big(\langle\mu_t^K,1\rangle-\langle\mu_t,1\rangle\big)^4\Big)\le C_T\big(I_4^4(K)+K^{-1}\big),\quad\forall t\in[0,T].$$

Proof. For the first bound on the moments of the total mass we refer to [12, Lemma 3.3]. To prove the convergence bound in the second part, we resort to algorithm (A) to represent the dynamics of the number of particles by the SDE

$$\begin{aligned}N_t^K &= N_0^K+\int_0^t\int\mathbf{1}_{\rho\le N_{s-}^K}\Big(\mathbf{1}_{\theta\le r}-\mathbf{1}_{r<\theta\le r+c\frac{N_{s-}^K}{K}}\Big)\,\mathcal{N}(\mathrm{d}s,\mathrm{d}\rho,\mathrm{d}\theta)\\ &= N_0^K+\int_0^t\Big(r-c\frac{N_s^K}{K}\Big)N_s^K\,\mathrm{d}s+M_t^K.\end{aligned}$$

Notice that the process $(M_t^K)_{t\ge0}$ is a martingale since, for all $t\ge0$,

$$\mathbb{E}\bigg(\int_0^t\int_{\mathbb{R}_+}\int_{\mathbb{R}_+}\bigg|\mathbf{1}_{\rho\le N_s^K}\Big(\mathbf{1}_{\theta\le r}-\mathbf{1}_{r<\theta\le r+c\frac{N_{s-}^K}{K}}\Big)\bigg|\,\mathrm{d}s\,\mathrm{d}\rho\,\mathrm{d}\theta\bigg)\le(r+c)\mathbb{E}\bigg(\int_0^t(N_s^K)^2\,\mathrm{d}s\bigg)<\infty,$$

by the previous part and the assumptions on the total mass of the system. The limiting mass in turn satisfies the dynamics

$$\langle\mu_t,1\rangle=\langle\mu_0,1\rangle+\int_0^t(r-c\langle\mu_s,1\rangle)\langle\mu_s,1\rangle\,\mathrm{d}s.$$

We will first establish an L^2 convergence bound for the total mass. Using Itô's formula we get

$$\Big(\frac{N_t^K}{K}-\langle\mu_t,1\rangle\Big)^2=\Big(\frac{N_0^K}{K}-\langle\mu_0,1\rangle\Big)^2+\int_0^t2\Big(\frac{N_{s-}^K}{K}-\langle\mu_{s-},1\rangle\Big)\,\mathrm{d}\Big(\frac{M_s^K}{K}\Big)$$

$$
\begin{aligned}
&+\int_0^t \left[2r\left(\frac{N_s^K}{K}-\langle \mu_s,1\rangle\right)^2-\left(\frac{N_s^K}{K}-\langle \mu_s,1\rangle\right)^2\left(\frac{N_s^K}{K}+\langle \mu_s,1\rangle\right)\right]\mathrm{d}s\\
&+\int_0^t\int \mathbf{1}_{\rho\le N_{s-}^K}\mathbf{1}_{r<\theta\le r+c\frac{N_{s-}^K}{K}}\left(\frac{1}{K}\right)^2\mathcal{N}(\mathrm{d}s,\mathrm{d}\rho,\mathrm{d}\theta)\\
&+\int_0^t\int \mathbf{1}_{\rho\le N_{s-}^K}\mathbf{1}_{\theta\le r}\left(\frac{1}{K}\right)^2\mathcal{N}(\mathrm{d}s,\mathrm{d}\rho,\mathrm{d}\theta).
\end{aligned}
$$

Bounding above the negative term in the second line by 0 gives us

$$
\begin{aligned}
\left(\frac{N_t^K}{K}-\langle \mu_t,1\rangle\right)^2 &\le \left(\frac{N_0^K}{K}-\langle \mu_0,1\rangle\right)^2+\int_0^t 2r\left(\frac{N_s^K}{K}-\langle \mu_s,1\rangle\right)^2\mathrm{d}s+\int_0^t \frac{r}{K}\left(\frac{N_s^K}{K}\right)\mathrm{d}s\\
&\quad+\int_0^t\frac{c}{K}\left(\frac{N_s^K}{K}\right)^2\mathrm{d}s+\int_0^t 2\left(\frac{N_{s-}^K}{K}-\langle \mu_{s-},1\rangle\right)\mathrm{d}\left(\frac{M_s^K}{K}\right)+\bar{M}_t^K+\tilde{M}_t^K\\
&\le \left(\frac{N_0^K}{K}-\langle \mu_0,1\rangle\right)^2+\int_0^t 2r\left(\frac{N_s^K}{K}-\langle \mu_s,1\rangle\right)^2\mathrm{d}s+\frac{rT}{K}\sup_{s\in[0,T]}\langle \mu_s^K,1\rangle\\
&\quad+\frac{cT}{K}\sup_{s\in[0,T]}\langle \mu_s^K,1\rangle^2+\int_0^t 2\left(\frac{N_{s-}^K}{K}-\langle \mu_{s-},1\rangle\right)\mathrm{d}\left(\frac{M_s^K}{K}\right)+\bar{M}_t^K+\tilde{M}_t^K,
\end{aligned}
\tag{5.1}
$$

where $(\bar{M}_t^K)_{t\ge 0}$ and $(\tilde{M}_t^K)_{t\ge 0}$ are compensated Poisson integrals. Let now $(\tau_m)_m$ be the sequence of stopping times defined by $\tau_m=\inf\{t>0:\overline{N}_t^K>m\}$ for $m\ge 1$ and $\tau_0=0$. Since $\overline{N}_s^K$ is increasing by one and $\overline{N}_{r-}^K\ge m+1=\overline{N}_{\tau_m}^K>\overline{N}_{s-}^K$ for all $r>\tau_m\ge s$, we have

$$
\begin{aligned}
\int_0^{t\wedge\tau_m} 2\left(\frac{N_{s-}^K}{K}-\langle \mu_{s-},1\rangle\right)\mathrm{d}\left(\frac{M_s^K}{K}\right)&=2\int_0^t \mathbf{1}_{\{\overline{N}_{s-}^K\le m\}}\left(\frac{N_{s-}^K}{K}-\langle \mu_{s-},1\rangle\right)\mathrm{d}\left(\frac{M_s^K}{K}\right)\\
&=2\int_0^t\int \phi(s,\rho,\theta)\tilde{\mathcal{N}}(\mathrm{d}s,\mathrm{d}\rho,\mathrm{d}\theta),
\end{aligned}
$$

with $\tilde{\mathcal{N}}$ the compensated measure associated with $\mathcal{N}$ and ϕ the predictable process

$$
\phi(s,\rho,\theta)=\mathbf{1}_{\overline{N}_{s-}^K\le m}\mathbf{1}_{\rho\le N_{s-}^K}\frac{1}{K}\left(\mathbf{1}_{\theta\le r}-\mathbf{1}_{r<\theta\le r+c\frac{N_{s-}^K}{K}}\right)\left(\frac{N_{s-}^K}{K}-\langle \mu_{s-},1\rangle\right).
$$

The inequality $N_s^K\le \overline{N}_s^K$ implies that

$$
\begin{aligned}
&\mathbb{E}\left(\int_0^t\int_0^\infty\int_0^\infty |\phi(s,\rho,\theta)|\,\mathrm{d}s\,\mathrm{d}\rho\,\mathrm{d}\theta\right)\\
&\qquad\le \mathbb{E}\left(\int_0^t \mathbf{1}_{\overline{N}_s^K\le m}(s)\frac{N_s^K}{K}\left(r+c\frac{N_s^K}{K}\right)\left(\frac{N_s^K}{K}+\langle \mu_s,1\rangle\right)\mathrm{d}s\right)\\
&\qquad\le \mathbb{E}\left(\int_0^t \frac{m}{K}\left(r+c\frac{m}{K}\right)\left(\frac{m}{K}+\langle \mu_s,1\rangle\right)\mathrm{d}s\right)\\
&\qquad\le C_{T,K,m}\left(1+\sup_{s\in[0,T]}\langle \mu_s,1\rangle\right),
\end{aligned}
$$

for each $t\in[0,T]$, and so the integral w.r.t. $\mathrm{d}\left(\frac{M_s^K}{K}\right)$ in (5.1) is a martingale. By similar reasonings, the stopped processes $(\bar{M}_{t\wedge\tau_m}^K)_{t\ge 0}$ and $(\tilde{M}_{t\wedge\tau_m}^K)_{t\ge 0}$ are also seen to be martingales. Taking expectation in (5.1) we get

$$
\mathbb{E}\left(\left(\frac{N_{t\wedge\tau_m}^K}{K}-\langle \mu_{t\wedge\tau_m},1\rangle\right)^2\right)
$$

$$
\begin{aligned}
&\le \mathbb{E}\bigg(\Big(\frac{N_0^K}{K} - \langle \mu_0, 1\rangle\Big)^2\bigg) + \mathbb{E}\bigg(\int_0^{t\wedge\tau_m} 2r\Big(\frac{N_s^K}{K} - \langle \mu_s, 1\rangle\Big)^2 \mathrm{d}s\bigg) + \frac{C_T}{K} \\
&\le \mathbb{E}\bigg(\Big(\frac{N_0^K}{K} - \langle \mu_0, 1\rangle\Big)^2\bigg) + \int_0^t 2r\mathbb{E}\bigg(\Big(\frac{N_{s\wedge\tau_m}^K}{K} - \langle \mu_{s\wedge\tau_m}, 1\rangle\Big)^2\bigg) \mathrm{d}s + \frac{C_T}{K},
\end{aligned}
$$

for all $t \in [0, T]$. Using Gronwall's lemma we obtain for all $t \in [0, T]$ that

$$
\mathbb{E}\bigg(\Big(\frac{N_{t\wedge\tau_m}^K}{K} - \langle \mu_{t\wedge\tau_m}, 1\rangle\Big)^2\bigg) \le \bigg(\mathbb{E}\bigg(\Big(\frac{N_0^K}{K} - \langle \mu_0, 1\rangle\Big)^2\bigg) + \frac{C_T}{K}\bigg) e^{2rT}. \tag{5.2}
$$

By Fatou's lemma, we then get $\mathbb{E}\Big(\big(\langle \mu_t^K, 1\rangle - \langle \mu_t, 1\rangle\big)^2\Big) \le C_T\big(I_2^2(K) + K^{-1}\big)$, but the bound (5.2) will be more practical for our purposes. Let us now address the L^4 bound. Applying Itô's formula again we get for all $t \ge 0$ that

$$
\begin{aligned}
\Big(\frac{N_t^K}{K} - \langle \mu_t, 1\rangle\Big)^4 = {}& \Big(\frac{N_0^K}{K} - \langle \mu_0, 1\rangle\Big)^4 + \int_0^t 4\Big(\frac{N_{s^-}^K}{K} - \langle \mu_{s^-}, 1\rangle\Big)^3 \mathrm{d}\Big(\frac{M_s^K}{K}\Big) \\
&+ \int_0^t \bigg[4r\Big(\frac{N_s^K}{K} - \langle \mu_s, 1\rangle\Big)^4 - 4\Big(\frac{N_s^K}{K} - \langle \mu_s, 1\rangle\Big)^4\Big(\frac{N_s^K}{K} + \langle \mu_s, 1\rangle\Big)\bigg] \mathrm{d}s \\
&+ \int_0^t \int \mathbf{1}_{\rho \le N_{s-}^K} \mathbf{1}_{\theta \le r} \bigg[\Big(\frac{N_{s-}^K}{K} - \langle \mu_{s-}, 1\rangle + \frac{1}{K}\Big)^4 \\
&- \Big(\frac{N_{s-}^K}{K} - \langle \mu_{s-}, 1\rangle\Big)^4 - 4\Big(\frac{N_{s-}^K}{K} - \langle \mu_{s-}, 1\rangle\Big)^3 \frac{1}{K}\bigg] \mathcal{N}(\mathrm{d}s, \mathrm{d}\rho, \mathrm{d}\theta) \\
&+ \int_0^t \int \mathbf{1}_{\rho \le N_{s-}^K} \mathbf{1}_{r < \theta \le r + c\frac{N_{s-}^K}{K}} \bigg[\Big(\frac{N_{s-}^K}{K} - \langle \mu_{s-}, 1\rangle - \frac{1}{K}\Big)^4 \\
&- \Big(\frac{N_{s-}^K}{K} - \langle \mu_{s-}, 1\rangle\Big)^4 + 4\Big(\frac{N_{s-}^K}{K} - \langle \mu_{s-}, 1\rangle\Big)^3 \frac{1}{K}\bigg] \mathcal{N}(\mathrm{d}s, \mathrm{d}\rho, \mathrm{d}\theta).
\end{aligned}
$$

Bounding the negative term in the second line by 0 and compensating the Poisson integrals gives us

$$
\begin{aligned}
&\Big(\frac{N_t^K}{K} - \langle \mu_t, 1\rangle\Big)^4 \\
&\quad \le \Big(\frac{N_0^K}{K} - \langle \mu_0, 1\rangle\Big)^4 + \int_0^t 4r\Big(\frac{N_s^K}{K} - \langle \mu_s, 1\rangle\Big)^4 \mathrm{d}s \\
&\qquad + \int_0^t rN_s^K\bigg(6\Big(\frac{N_s^K}{K} - \langle \mu_s, 1\rangle\Big)^2 \frac{1}{K^2} + 4\Big(\frac{N_s^K}{K} - \langle \mu_s, 1\rangle\Big)\frac{1}{K^3} + \frac{1}{K^4}\bigg) \mathrm{d}s \\
&\qquad + \int_0^t cN_s^K \frac{N_s^K}{K}\bigg(6\Big(\frac{N_s^K}{K} - \langle \mu_s, 1\rangle\Big)^2 \frac{1}{K^2} - 4\Big(\frac{N_s^K}{K} - \langle \mu_s, 1\rangle\Big)\frac{1}{K^3} + \frac{1}{K^4}\bigg) \mathrm{d}s \\
&\qquad + \int_0^t 4\Big(\frac{N_{s^-}^K}{K} - \langle \mu_{s^-}, 1\rangle\Big)^3 \mathrm{d}\Big(\frac{M_s^K}{K}\Big) + R_t^K + \bar{R}_t^K,
\end{aligned}
$$

for all $t \ge 0$, where $(R_t^K)_{t\ge 0}$ and $(\bar{R}_t^K)_{t\ge 0}$ are compensated Poisson integrals. Using Young's inequality we deduce for all $t \in [0, T]$ that

$$
\begin{aligned}
&\Big(\frac{N_t^K}{K} - \langle \mu_t, 1\rangle\Big)^4 \\
&\quad \le \Big(\frac{N_0^K}{K} - \langle \mu_0, 1\rangle\Big)^4 + C\int_0^t \Big(\frac{N_s^K}{K} - \langle \mu_s, 1\rangle\Big)^4 \mathrm{d}s + \frac{C}{K^2}\int_0^t \Big(\frac{N_s^K}{K} - \langle \mu_s, 1\rangle\Big)^2 \mathrm{d}s
\end{aligned}
$$

$$+\frac{C_T}{K^3}\sup_{s\in[0,T]}\langle\mu_s^K,1\rangle+\frac{C_T}{K}\sup_{s\in[0,T]}\langle\mu_s^K,1\rangle^2+\frac{C_T}{K}\sup_{s\in[0,T]}\langle\mu_s^K,1\rangle^4$$
$$+\int_0^t 4\Big(\frac{N_{s^-}^K}{K}-\langle\mu_{s^-},1\rangle\Big)^3\mathrm{d}\Big(\frac{M_s^K}{K}\Big)+R_t^K+\bar{R}_t^K. \tag{5.3}$$

Proceeding in a similar way as in the proof of the bound (5.2), we can verify again that the three processes in the last line are martingales if stopped at $\tau_m=\inf\{t>0:\overline{N}_t^K>m\}$. Thus, stopping the inequality (5.3) and taking expectation yields for all $t\in[0,T]$ that

$$\begin{aligned}\mathbb{E}\bigg(\Big(\frac{N_{t\wedge\tau_m}^K}{K}-\langle\mu_{t\wedge\tau_m},1\rangle\Big)^4\bigg)&\le I_4^4(K)+\frac{C_T}{K}+C\int_0^t\mathbb{E}\bigg(\Big(\frac{N_{s\wedge\tau_m}^K}{K}-\langle\mu_{s\wedge\tau_m},1\rangle\Big)^4\bigg)\mathrm{d}s\\&\quad+\frac{C}{K^2}\int_0^t\mathbb{E}\bigg(\Big(\frac{N_{s\wedge\tau_m}^K}{K}-\langle\mu_{s\wedge\tau_m},1\rangle\Big)^2\bigg)\mathrm{d}s\\&\le I_4^4(K)+\frac{C_T}{K}+C\int_0^t\mathbb{E}\bigg(\Big(\frac{N_{s\wedge\tau_m}^K}{K}-\langle\mu_{s\wedge\tau_m},1\rangle\Big)^4\bigg)\mathrm{d}s\\&\quad+\frac{C_TT}{K^2}\Big(I_2^2(K)+\frac{1}{K}\Big),\end{aligned}$$

where we used (5.2) to obtain the second inequality. Gronwall's inequality and then Fatou's lemma yield at last

$$\mathbb{E}\bigg(\Big(\frac{N_t^K}{K}-\langle\mu_t,1\rangle\Big)^4\bigg)\le I_4^4(K)+C_T\Big(\frac{1}{K}+\frac{I_2^2(K)}{K^2}\Big),\quad\forall t\in[0,T],$$

and we obtain the asserted bound noting that $I_2^2(K)\le\sqrt{I_4^4(K)}\le 1+I_4^4(K)$. □

We prove now Proposition 11, which relates the solution $(\mu_t)_{t\ge0}$ of equation (1.2) to a nonlinear process of McKean-Vlasov type.

Proof of Proposition 11. Pathwise existence and uniqueness for the SDE (3.1) comes from the fact that the (non-homogeneous) coefficients are Lipschitz functions of Y_s, thanks to (H.2). In order to characterize the flow of time-marginal laws of $(Y_t)_{t\ge0}$, consider a function $f\in C^{1,2}([0,T]\times\mathbb{R}^d)$ satisfying the conditions in Theorem 2. By Itô's formula we obtain

$$\begin{aligned}f(t,Y_t)&=f(0,Y_0)+\int_0^t\frac{\partial f(s,Y_s)}{\partial s}\mathrm{d}s+\int_0^t\nabla f(s,Y_s)^{\mathrm{t}}b(Y_s,H*\mu_s(Y_s))\,\mathrm{d}s\\&\quad+\int_0^t\nabla f(s,Y_s)^{\mathrm{t}}\sigma(Y_s,G*\mu_s(Y_s))\,\mathrm{d}W_s+\frac12\int_0^t\mathrm{Tr}(a(Y_s,G*\mu_s(Y_s))\mathrm{Hess}f(s,Y_s))\,\mathrm{d}s.\end{aligned}$$

Taking expectation shows that the law of the time-marginal is a weak solution of equation (3.2) with respect to that set of test functions. Now, consider the function $h(t,x)=\langle\mu_t,1\rangle f(t,x)$. By equation (3.2) we get for all $t\ge0$ that

$$\begin{aligned}&\langle\tilde\mu_t,h(t,\cdot)\rangle\\&\quad=\langle\tilde\mu_0,h(0,\cdot)\rangle+\int_0^t\langle\tilde\mu_s,\partial_sh(s,\cdot)+L_{\mu_s}h(s,\cdot)\rangle\,\mathrm{d}s\\&\quad=\langle\langle\mu_0,1\rangle\tilde\mu_0,f(0,\cdot)\rangle+\int_0^t\langle\tilde\mu_s,f(s,\cdot)\partial_s\langle\mu_s,1\rangle+\langle\mu_s,1\rangle\partial_sf(s,\cdot)+\langle\mu_s,1\rangle L_{\mu_s}f(s,\cdot)\rangle\,\mathrm{d}s\\&\quad=\langle\langle\mu_0,1\rangle\tilde\mu_0,f(0,\cdot)\rangle+\int_0^t\langle\langle\mu_s,1\rangle\tilde\mu_s,\partial_sf(s,\cdot)+L_{\mu_s}f(s,\cdot)+(r-c\langle\mu_s,1\rangle)f(s,\cdot)\rangle\,\mathrm{d}s,\end{aligned}$$

which implies that $(\xi_t)_{t\geq 0} \coloneqq (\langle \mu_t, 1\rangle \tilde{\mu}_t)_{t\geq 0}$ satisfies the following “linearized” version of equation (1.2):

$$\langle \xi_t, f(t,\cdot)\rangle = \langle \mu_0, f(0,\cdot)\rangle + \int_0^t \langle \xi_s, \partial_s f(s,\cdot) + L_{\mu_s} f(s,\cdot) + (r - c\langle \mu_s, 1\rangle) f(s,\cdot)\rangle \, \mathrm{d}s, \quad (5.4)$$

where $(\mu_t)_{t\geq 0}$ is the given solution to equation (1.2). With similar (indeed simpler) arguments as in the uniqueness part of Theorem 2 (see [12, Section 4]) one can show that uniqueness of weak solutions (with respect to the same class of test functions) of equation (5.4) holds. But $(\xi_t)_{t\geq 0} = (\mu_t)_{t\geq 0}$ is also solution of the linear equation (5.4), because $(\mu_t)_{t\geq 0}$ solves (1.2). We deduce that $\langle \mu_t, 1\rangle \tilde{\mu}_t = \mu_t$ for all $t \geq 0$, and in particular that $\tilde{\mu}_t = \bar{\mu}_t$.

The previous identity yields $\langle \mu_t, f\rangle = \mathbb{E}(\langle \mu_t, 1\rangle f(Y_t))$ for every bounded measurable f, and the fact that $(\langle \mu_t, 1\rangle)_{t\geq 0}$ is the unique solution of equation (3.3) is readily obtained by taking $f = 1$ in Theorem 2, recalling also that the local Lipschitz character of the ODE's coefficient ensures uniqueness for it. □

The following propagation of moments result for the unique solution of equation (3.2) will be needed.

Lemma 23. Assume (H) and let $(\mu_t)_{t\geq 0}$ be the unique solution in $\mathcal{M}^+(\mathbb{R}^d)$ of (1.2). Letting $(\bar{\mu}_t)_{t\geq 0}$ be the normalizations of $(\mu_t)_{t\geq 0}$, for each $T > 0$ and $q \geq 2$ there is a constant $C'_T > 0$ such that

$$\sup_{t\in[0,T]} M_q(\bar{\mu}_t) < C'_T(1 + M_q(\bar{\mu}_0)).$$

Proof. We will use the fact that diffusion process $(Y_t)_{t\geq 0}$ studied in Proposition 11 satisfies $\mathbb{E}(\|Y_t\|^q) = M_q(\bar{\mu}_t)$. Applying Itô's formula to $\|Y_t\|^q, t \geq 0$ with $q \geq 2$ yields

$$\begin{aligned}
\|Y_t\|^q = \|Y_0\|^q &+ \int_0^t q\|Y_s\|^{q-2} Y_s^{\mathrm{t}} b(Y_s, H * \mu_s(Y_s)) \, \mathrm{d}s + \int_0^t q\|Y_s\|^{q-2} Y_s^{\mathrm{t}} \sigma(Y_s, G * \mu_s(Y_s)) \, \mathrm{d}B_s \\
&+ \frac{1}{2}\sum_{i,j=1}^d \sum_{k=1}^d \int_0^t \Big(q(q-2)\|Y_s\|^{q-4} |Y_s^{(i)}||Y_s^{(j)}| + \delta_{ij}\|Y_s\|^{q-2}\Big) \\
&\qquad \times \sigma^{(ik)}(Y_s, G * \mu_s(Y_s)) \sigma^{(jk)}(Y_s, G * \mu_s(Y_s))) \, \mathrm{d}s. \qquad (5.5)
\end{aligned}$$

Since b is Lipschitz we have $\|b(Y_s, H * \mu_s(Y_s))\| \leq C\big(1 + \|Y_s\| + |H * \mu_s(Y_s)|\big)$ with $|H * \mu_s(Y_s)| = |\int H(x - Y_s)\mu_s(\mathrm{d}x)| \leq \|H\|_\infty \sup_{t\in[0,T]} |\langle \mu_t, 1\rangle|$ and similarly for σ and G. We thus get that

$$\|b(Y_s, H * \mu_s(Y_s))\| \leq C_T\big(1 + \|X_s\|\big) \text{ and } \|\sigma(X_s, G * \mu_s(X_s))\| \leq C_T\big(1 + \|X_s\|\big).$$

for all $s \in [0, T]$ and some constant $C_T > 0$. Using this in (5.5) gives us the bound

$$\begin{aligned}
\|Y_t\|^q \leq \|Y_0\|^q &+ C\int_0^t \|Y_s\|^{q-2} \, \mathrm{d}s + C\int_0^t \|Y_s\|^{q-1} \, \mathrm{d}s + C\int_0^t \|Y_s\|^q \, \mathrm{d}s \\
&+ \int_0^t q\|Y_s\|^{q-2} Y_s^{\mathrm{t}} \sigma(Y_s, G * \mu_s(Y_s)) \, \mathrm{d}B_s \,, \quad \forall t \in [0, T].
\end{aligned}$$

Let now $(\tau_n)_{n\in\mathbb{N}}$ be a localizing sequence for the local martingale on the right hand side. Taking expectation of the stopped process yields, for all $t \in [0, T]$, that

$$\mathbb{E}(\|Y_{t\wedge\tau_n}\|^q) \leq \mathbb{E}(\|Y_0\|^q) + C\int_0^t \mathbb{E}(\|Y_{s\wedge\tau_n}\|^{q-2}) \, \mathrm{d}s + C\int_0^t \mathbb{E}(\|Y_{s\wedge\tau_n}\|^{q-1}) \, \mathrm{d}s$$

$$+ C \int_0^t \mathbb{E}(\|Y_{s\wedge\tau_n}\|^q)\,\mathrm{d}s.$$

Notice that by Hölder's inequality, one gets

$$\int_0^t \mathbb{E}(\|Y_{s\wedge\tau_n}\|^{q-1})\,\mathrm{d}s \le \int_0^t \mathbb{E}(\|Y_{s\wedge\tau_n}\|^q)^{\frac{q-1}{q}}\,\mathrm{d}s \le C_T + C\int_0^t \mathbb{E}(\|Y_{s\wedge\tau_n}\|^q)\,\mathrm{d}s,$$

for all $t \in [0,T]$, and a similar bound holds for the term of order $q-2$. Combined with the previous, this entails

$$\mathbb{E}(\|Y_{t\wedge\tau_n}\|^q) \le \mathbb{E}(\|Y_0\|)^q + C_T + C\int_0^t \mathbb{E}(\|Y_{s\wedge\tau_n}\|^q)\,\mathrm{d}s,$$

from where Gronwall's lemma yields, for all $t \in [0,T]$,

$$\mathbb{E}(\|Y_{t\wedge\tau_n}\|^q) \le C_T(\mathbb{E}(\|Y_0\|)^q + 1).$$

We conclude with Fatou's lemma taking $n \to \infty$. □

In order to check that condition (C.3) holds, we need some additional bounds stated in the next two results (respectively analogous to Lemmas 20 and 19 in the pure branching case). In particular, the following result will be used to control the joint evolution of coupled particles in the two systems, between birth and death events.

Lemma 24. Assume (H.2) and (H.3). Let N and $K \in \mathbb{N}\setminus\{0\}$ be deterministic and fixed, and consider the diffusion processes $(X^n)_{n=1}^N$ in $(\mathbb{R}^d)^N$ evolving according to

$$\mathrm{d}X_t^n = b(X_t^n, H * \xi_t^K(X_t^n))\,\mathrm{d}t + \sigma(X_t^n, G*\xi_t^K(X_t^n))\,\mathrm{d}B_t^n, \quad t \ge 0,$$

where $(B^n)_{n=1}^N$ are independent Brownian motions in $\mathbb{R}^d$ and ξ_t^K stands for the empirical measure $\xi_t^K = \frac{1}{K}\sum_{n=1}^N \delta_{X_t^n}$ of constant mass N/K. Consider also N i.i.d. copies $(Y^n)_{n=1}^N$ of the process (3.1),

$$\mathrm{d}Y_t^n = b(Y_t^n, H*\mu_t(Y_t^n))\,\mathrm{d}t + \sigma(Y_t^n, G*\mu_t(Y_t^n))\,\mathrm{d}B_t^n, \quad t \ge 0,$$

driven by the same Brownian motions $(B^n)_{n=1}^N$ and where $(\mu_t)_{t\ge0}$ is the unique solution of (1.2). For each $T > 0$, there is $C_T > 0$ not depending on K nor on N such that for all $0 < u < t < T$ and each $n = 1,\dots,N$,

$$\mathbb{E}(\|X_t^n - Y_t^n\|^2 - \|X_u^n - Y_u^n\|^2) \le C_T\int_u^t \mathbb{E}(\|X_s^n - Y_s^n\|^2)\,\mathrm{d}s + \int_u^t \mathbb{E}\Big(\big\|\xi_s^K - \mu_s\big\|_{\mathrm{BL}^*}^2\Big)\,\mathrm{d}s.$$

Proof. We first check that the running supremum of each process (X^n) is square integrable. Using similar bounds as in the proof of Lemma 23, we get for each $t \in [0,T]$,

$$\begin{aligned}
\|X_t^n\|^2 &\le \|X_0^n\|^2 + \int_0^t 2\|X_s^n\|\|b(X_s^n, H*\xi_s^K(X_s^n))\|\,\mathrm{d}s + \int_0^t 2(X_s^n)^{\mathrm{t}}\sigma(X_s^n, G*\xi_s^K(X_s^n))\,\mathrm{d}B_s \\
&\quad + \int_0^t \|\sigma(X_s^n, G*\xi_s^K(X_s^n))\|^2\,\mathrm{d}s \\
&\le \|X_0^n\|^2 + C_T + C\int_0^t \|X_s^n\|\,\mathrm{d}s + C\int_0^t \|X_s^n\|^2\,\mathrm{d}s + C\int_0^t \|X_s^n\||H*\xi_s^K(X_s^n)|\,\mathrm{d}s \\
&\quad + C\int_0^t |G*\xi_s^K(X_s^n)|^2\,\mathrm{d}s + \int_0^t 2(X_s^n)^{\mathrm{t}}\sigma(X_s^n, G*\xi_s^K(X_s^n))\,\mathrm{d}B_s \\
&\le \|X_0^n\|^2 + C_T + CT\|H\|_\infty^2\left(\frac{N}{K}\right)^2 + CT\|G\|_\infty^2\left(\frac{N}{K}\right)^2 + C\int_0^t \|X_s^n\|^2\,\mathrm{d}s \\
&\quad + \int_0^t 2(X_s^n)^{\mathrm{t}}\sigma(X_s^n, G*\xi_s^K(X_s^n))\,\mathrm{d}B_s,
\end{aligned}$$

since, in the present lemma's setting, $\langle \xi^K_s, 1\rangle = N/K$ for all $s \geq 0$. Let $(\tau_m)_{m\in\mathbb{N}}$ be a localizing sequence for the local martingale in the previous inequality. As in the proof of Lemma 19 we localize and then we take supremum until time $t \wedge \tau_m$ on both sides, obtaining that

$$\begin{aligned}\sup_{u\in[0,t\wedge\tau_m]} \|X^n_u\|^2 \leq \|X^n_0\|^2 + C_T + CT\|H\|^2_\infty \left(\frac{N}{K}\right)^2 + CT\|G\|^2_\infty \left(\frac{N}{K}\right)^2 \\ + C\int_0^t \sup_{u\in[0,s\wedge\tau_m]} \|X^n_u\|^2 \,\mathrm{d}s \\ + \sum_{i,j=1}^d \left(\sup_{u\in[0,t\wedge\tau_m]} \left| \int_0^u 2(X^n_s)^{(i)} \sigma^{(ij)}(X^n_s, G * \xi^K_s(X^n_s))\,\mathrm{d}B^{(j)}_s \right|\right).\end{aligned}$$

The expectation of the last term is controlled using the BDG inequality by

$$\begin{aligned}\sum_{i,j=1}^d \mathbb{E}\left(\sup_{u\in[0,t\wedge\tau_m]} \left| \int_0^u 2(X^n_s)^{(i)} \sigma^{(ij)}(X^n_s, G * \xi^K_s(X^n_s))\,\mathrm{d}B^{(j)}_s \right|\right) \\ \leq \sum_{i,j=1}^d \mathbb{E}\left(\left(\int_0^{t\wedge\tau_m} 4\Big((X^n_s)^{(i)}\sigma^{(ij)}(X^n_s, G*\xi^K_s(X^n_s))\Big)^2 \mathrm{d}s\right)^{\frac{1}{2}}\right) \\ \leq C\mathbb{E}\left(\left(\int_0^{t\wedge\tau_m} \|X^n_s\|^2 \|\sigma(X^n_s, G*\xi^K_s(X^n_s))\|^2 \,\mathrm{d}s\right)^{\frac{1}{2}}\right) \\ \leq C\mathbb{E}\left(\left(1 + \|G\|^2_\infty \left(\frac{N}{K}\right)^2\right)^{\frac{1}{2}} \left(\int_0^t \|X^n_{s\wedge\tau_m}\|^2 \,\mathrm{d}s\right)^{\frac{1}{2}}\right) \\ \leq \left(1 + \left(\frac{N}{K}\right)^2\right)\left(C_T + C_T\int_0^t \mathbb{E}(\|X^n_{s\wedge\tau_m}\|^2)\,\mathrm{d}s\right).\end{aligned}$$

This allows us to deduce for all $t \in [0,T]$ that

$$\mathbb{E}\left(\sup_{u\in[0,t\wedge\tau_m]} \|X^n_u\|^2\right) \leq \mathbb{E}\big(\|X^n_0\|^2\big) + C_{T,N,K} + C_{T,N,K}\int_0^t \mathbb{E}\left(\sup_{u\in[0,s\wedge\tau_m]} \|X^n_u\|^2\right)\mathrm{d}s,$$

where $C_{T,N,K}$ is a constant depending on T, N and K (recall N and K are deterministic in the setting of this lemma). From this last inequality, Gronwall's lemma and monotone convergence when $m \to \infty$ yield

$$\mathbb{E}\left(\sup_{t\in[0,T]} \|X^n_t\|^2\right) < \infty.$$

A similar argument can be applied to the process $(Y^n_t)_{t\geq 0}$ in order to obtain the same conclusion. We now apply Itô's formula for each fixed n to get

$$\begin{aligned}\|X^n_t - Y^n_t\|^2 = \|X^n_u - Y^n_u\|^2 + \int_u^t 2(X^n_s - Y^n_s)^{\mathrm{t}}\big(b(X^n_s, H*\xi^K_s(X^n_s)) - b(Y^n_s, H*\mu_s(Y^n_s))\big)\,\mathrm{d}s \\ + \int_u^t 2(X^n_s - Y^n_s)^{\mathrm{t}}\big(\sigma(X^n_s, G*\xi^K_s(X^n_s)) - \sigma(Y^n_s, G*\mu_s(Y^n_s))\big)\,\mathrm{d}B^n_s \\ + \sum_{i,j=1}^d \int_u^t \big(\sigma^{(ij)}(X^n_s, G*\xi^K_s(X^n_s)) - \sigma^{(ij)}(Y^n_s, G*\mu_s(Y^n_s))\big)^2\,\mathrm{d}s.\end{aligned}$$

The Lipschitz character of the coefficients granted by (H.2) imply the bound

$$\|X^n_t - Y^n_t\|^2 \leq \|X^n_u - Y^n_u\|^2$$

$$
\begin{aligned}
&+ C\int_u^t \left(\|X_s^n - Y_s^n\|^2 + \|X_s^n - Y_s^n\||H * \xi_s^K(X_s^n) - H * \mu_s(Y_s^n)|\right) \mathrm{d}s \\
&+ C\int_u^v \left(\|X_s^n - Y_s^n\|^2 + |G * \xi_s^K(X_s^n) - G * \mu_s(Y_s^n)|^2\right) \mathrm{d}s \\
&+ \int_u^t 2(X_s^n - Y_s^n)^{\mathrm{t}}(\sigma(X_s^n, G * \xi_s^K(X_s^n)) - \sigma(Y_s^n, G * \mu_s(Y_s^n)))\, \mathrm{d}B_s^n.
\end{aligned}
$$

Assumption (H.3) implies that the function $H(\cdot - x)$ is bounded and Lipschitz for each $x \in \mathbb{R}^d$, hence

$$
\begin{aligned}
\left|H * \xi_s^K(X_s^n) - H * \mu_s(Y_s^n)\right| &\leq \left|H * \xi_s^K(X_s^n) - H * \mu_s(X_s^n)\right| + |H * \mu_s(X_s^n) - H * \mu_s(Y_s^n)| \\
&\leq C\|\xi_s^K - \mu_s\|_{\mathrm{BL}^*} + C\|\mu_s\|_{\mathrm{BL}^*}\|X_s^n - Y_s^n\|,
\end{aligned}
$$

holds for all $s \geq 0$, and similarly for the terms involving G. Then, the uniform bound on the mass of $(\mu_t)_{t\geq 0}$ on finite time intervals allows us to deduce for all $0 < u < t < T$ that

$$
\begin{aligned}
\|X_t^n - Y_t^n\|^2 &\leq \|X_u^n - Y_u^n\|^2 + C\int_u^t \left(\|X_s^n - Y_s^n\|^2 + \|X_s^n - Y_s^n\|\|\xi_s^K - \mu_s\|_{\mathrm{BL}^*}\right) \mathrm{d}s \\
&\quad + C\int_u^v \left(\|X_s^n - Y_s^n\|^2 + \|\xi_s^K - \mu_s\|_{\mathrm{BL}^*}^2\right) \mathrm{d}s \\
&\quad + \int_u^t 2(X_s^n - Y_s^n)^{\mathrm{t}}(\sigma(X_s^n, G * \xi_s^K(X_s^n)) - \sigma(Y_s^n, G * \mu_s(Y_s^n)))\, \mathrm{d}B_s^n \\
&\leq \|X_u^n - Y_u^n\|^2 + C\int_u^t \left(\|X_s^n - Y_s^n\|^2 + \|\xi_s^K - \mu_s\|_{\mathrm{BL}^*}^2\right) \mathrm{d}s \\
&\quad + \int_u^t 2(X_s^n - Y_s^n)^{\mathrm{t}}(\sigma(X_s^n, G * \xi_s^K(X_s^n)) - \sigma(Y_s^n, G * \mu_s(Y_s^n)))\, \mathrm{d}B_s^n,
\end{aligned}
$$

where we used Young's inequality for the second inequality, and where C is a constant depending on $T > 0$ but not on K nor N that changed from line to line. By considering a localizing sequence $(\tau_m)_m$ for the local martingale on the right hand side, we can take expectation of the stopped processes to obtain

$$
\begin{aligned}
\mathbb{E}(\|X_{t\wedge\tau_m}^n - Y_{t\wedge\tau_m}^n\|^2) &\leq \mathbb{E}(\|X_u^n - Y_u^n\|^2) + C\int_u^t \mathbb{E}(\|X_{s\wedge\tau_m}^n - Y_{s\wedge\tau_m}^n\|^2)\, \mathrm{d}s \\
&\quad + \int_u^t \mathbb{E}\left(\left\|\mu_{s\wedge\tau_m}^K - \mu_{s\wedge\tau_m}\right\|_{\mathrm{BL}^*}^2\right) \mathrm{d}s,
\end{aligned}
$$

for all $0 < u < t < T$. Thanks to the second moments controls on the running suprema of X^n and Y^n, and since the total mass of μ_t^K is constant in the context of the present lemma, we can use dominated convergence to take $m \to \infty$ and conclude the proof. $\square$

The following bound gathering all the previous estimates will allow us to check that condition (C.3) holds.

Lemma 25. Assume (H) and let $(\mu_t)_{t\geq 0}$ be the unique solution in $\mathcal{M}^+(\mathbb{R}^d)$ of (1.2) and $(\bar{\mu}_t)_{t\geq 0}$ its corresponding normalization. Consider $(Y_t^n)_{t\geq 0}, (X_t^n)_{t\geq 0}$, and $(\nu_t^K)_{t\geq 0}$ as constructed in algorithm (A). Then, for all $t \in [0, T]$ we have

$$
\mathbb{E}\Big(\frac{1}{K}\sum_{n=1}^{N_t^K} \|X_t^n - Y_t^n\|^2\Big) \leq C_T\left[I_4^2(K) + K^{-\frac{1}{2}} + \int_0^T \mathbb{E}\Big(\frac{N_s^K}{K} W_2^2(\bar{\nu}_s^K, \bar{\mu}_s)\Big)\, \mathrm{d}s\right],
$$

where $C_T > 0$ is a constant that depends on T and on the parameters of the model.

Proof. As in the proof of Lemma 20 we consider the product empirical measure $\eta_t^K := \frac{1}{K}\sum_{n=1}^{N_t^K}\delta_{(X_t^n,Y_t^n)}$ and decompose again

$$\mathbb{E}\Big(\frac{1}{K}\sum_{n=1}^{N_t^K}|X_t^n - Y_t^n|^2\Big) = \mathbb{E}(\langle \eta_t^K, d_2\rangle),$$

in terms of the sequence of jump times $(T_m)_{m\in\mathbb{N}}$, as in (4.5). We can proceed in a similar way as in (4.6) to control the evolution between jumps, now with help of Lemma 24, and control the contributions in the jump instants in the same way as in (4.8), to obtain for all $t \in [0,T]$ that

$$\begin{aligned}\mathbb{E}(\langle \eta_t^K, d_2\rangle) \le C\int_0^t \mathbb{E}(\langle \eta_s^K, d_2\rangle)\,\mathrm{d}s + C\int_0^t \mathbb{E}\Big(\frac{N_s^K}{K}W_2^2(\bar\nu_s^K,\bar\mu_s)\Big)\,\mathrm{d}s \\ + C\int_0^t \mathbb{E}\Big(\frac{N_s^K}{K}\|\mu_s^K - \mu_s\|_{\mathrm{BL}^*}^2\Big)\,\mathrm{d}s,\end{aligned}$$

where C is a positive constant depending on T. We observe that, compared to the case dealt with in the previous section, the interaction at the level of the dynamics only results in the addition of the last term in the previous inequality. In order to bound this new term, we use Lemma 4 to get

$$\begin{aligned}&\mathbb{E}\Big(\frac{N_s^K}{K}\|\mu_s^K - \mu_s\|_{\mathrm{BL}^*}^2\Big) \\ &\qquad\le \mathbb{E}\Big(\frac{N_s^K}{K}\Big(\langle\mu_s,1\rangle\|\bar\mu_s^K - \bar\mu_s\|_{\mathrm{BL}^*} + \Big|\frac{N_s^K}{K} - \langle\mu_s,1\rangle\Big|\Big)^2\Big) \\ &\qquad\le 2\sup_{u\in[0,T]}\langle\mu_u,1\rangle^2\mathbb{E}\Big(\frac{N_s^K}{K}\|\bar\mu_s^K - \bar\mu_s\|_{\mathrm{BL}^*}^2\Big) + 2\mathbb{E}\Big(\frac{N_s^K}{K}\Big|\frac{N_s^K}{K} - \langle\mu_s,1\rangle\Big|^2\Big) \\ &\qquad\le C\mathbb{E}\Big(\frac{N_s^K}{K}\|\bar\mu_s^K - \bar\nu_s^K\|_{\mathrm{BL}^*}^2\Big) + C\mathbb{E}\Big(\frac{N_s^K}{K}\|\bar\mu_s - \bar\nu_s^K\|_{\mathrm{BL}^*}^2\Big) \\ &\qquad\quad + 2\mathbb{E}\Big(\frac{N_s^K}{K}\Big|\frac{N_s^K}{K} - \langle\mu_s,1\rangle\Big|^2\Big), \quad s\in[0,T],\end{aligned}$$

using also the uniform bounds for the mass of the solution to equation (11) on finite time intervals. To control the first term of the right hand side, we relate it to the Wasserstein distance using again Lemma 4, which yields

$$\mathbb{E}\Big(\frac{N_s^K}{K}\|\bar\mu_s^K - \bar\nu_s^K\|_{\mathrm{BL}^*}^2\Big) \le \mathbb{E}\Big(\frac{N_s^K}{K}W_2^2(\bar\mu_s^K,\bar\nu_s^K)\Big) \le \mathbb{E}(\langle\eta_s^K,d_2\rangle), \quad \forall s\ge 0.$$

We do the same with the second term to get

$$\mathbb{E}\Big(\frac{N_s^K}{K}\|\bar\mu_s - \bar\nu_s^K\|_{\mathrm{BL}^*}^2\Big) \le \mathbb{E}\Big(\frac{N_s^K}{K}W_2^2(\bar\mu_s,\bar\nu_s^K)\Big).$$

We thus obtain the inequality

$$\begin{aligned}\mathbb{E}(\langle \eta_t^K, d_2\rangle) \le C\int_0^t \mathbb{E}(\langle \eta_s^K, d_2\rangle)\,\mathrm{d}s + C\int_0^t \mathbb{E}\Big(\frac{N_s^K}{K}W_2^2(\bar\nu_s^K,\bar\mu_s)\Big)\,\mathrm{d}s \\ + 2\int_0^t \mathbb{E}\Big(\frac{N_s^K}{K}\Big|\frac{N_s^K}{K} - \langle\mu_s,1\rangle\Big|^2\Big)\,\mathrm{d}s,\end{aligned}$$

for all $s \in [0,T]$, where only the last term needs to be controlled. Using Hölder's inequality yields

$$\mathbb{E}\Big(\frac{N_s^K}{K}\Big|\frac{N_s^K}{K} - \langle\mu_s,1\rangle\Big|^2\Big) \le \mathbb{E}\Big(\Big(\frac{N_s^K}{K}\Big)^2\Big)^{\frac12}\mathbb{E}\Big(\Big|\frac{N_s^K}{K} - \langle\mu_s,1\rangle\Big|^4\Big)^{\frac12},$$

where the first factor on the r.h.s. is controlled by Lemma 22. Thanks to the second bound in Lemma 22, we obtain that

$$\mathbb{E}(\langle \eta_t^K, d_2\rangle) \le C\int_0^t \mathbb{E}(\langle \eta_s^K, d_2\rangle)\,\mathrm{d}s + C\int_0^t \mathbb{E}\Big(\frac{N_s^K}{K} W_2^2(\bar\nu_s^K,\bar\mu_s)\Big)\,\mathrm{d}s + C_T\Big(I_4^2(K)+\frac{1}{\sqrt{K}}\Big),$$

for each $t\in[0,T]$. Finally, Gronwall's lemma yields

$$\mathbb{E}(\langle \eta_t^K, d_2\rangle) \le C_T\Big[I_4^2(K)+\frac{1}{\sqrt{K}}+\int_0^T \mathbb{E}\Big(\frac{N_s^K}{K} W_2^2(\bar\nu_s^K,\bar\mu_s)\Big)\,\mathrm{d}s\Big]e^{CT}, \quad \forall t\in[0,T]. \qquad \square$$

We deduce the following result.

Corollary 26. Assume (H). Then, condition (C.3) holds for the systems $(\mu_t^K)_{t\ge 0}$ and $(\nu_t^K)_{t\ge 0}$ constructed in algorithm (A).

Proof. Applying Lemma 25, Lemma 6 and noting that $1/\sqrt{K}\le CR_{d,q}(K)$, we obtain the bound

$$\mathbb{E}(\langle \eta_t^K, d_2\rangle) \le C_T\Big(I_4^2(K)+R_{d,q}(K)\Big), \quad \forall t\in[0,T]. \tag{5.6}$$

It suffices to combine this with the inequality

$$\mathbb{E}\big(\frac{N_t^K}{K}W_2^2\big(\bar\mu_t^K,\bar\nu_t^K\big)\big) \le \mathbb{E}\big(\frac{1}{K}\sum_{n=1}^{N_t^K}\|X_t^n-Y_t^n\|^2\big),$$

to conclude. $\square$

Finally, everything is in place to prove the main result.

Proof of Theorem 3 under (H). Following (4.10) and using Lemma 22, we get

$$\begin{aligned}
&\mathbb{E}\left(\|\mu_t^K-\mu_t\|_{\mathrm{BL}^*}\right)\\
&\qquad \le \left(\mathbb{E}\Big(\frac{N_t^K}{K}W_2^2\big(\bar\nu_t^K,\bar\mu_t\big)\Big)^{\frac12} + \mathbb{E}\Big(\frac{N_t^K}{K}W_2^2\big(\bar\nu_t^K,\bar\mu_t^K\big)\Big)^{\frac12}\right)\mathbb{E}\Big(\frac{N_t^K}{K}\Big)^{\frac12}\\
&\qquad\quad + \mathbb{E}\Big(\big(\langle\mu_t^K,1\rangle-\langle\mu_t,1\rangle\big)^2\Big)^{\frac12}\\
&\qquad \le C_T\left(\mathbb{E}\Big(\frac{N_t^K}{K}W_2^2\big(\bar\nu_t^K,\bar\mu_t\big)\Big)^{\frac12} + \mathbb{E}\Big(\frac{N_t^K}{K}W_2^2\big(\bar\nu_t^K,\bar\mu_t^K\big)\Big)^{\frac12} + I_2(K)+K^{-1/2}\right),
\end{aligned}$$

for all $t\in[0,T]$. As in the previous section, thanks to condition (C), Lemma 22, Lemma 23, and Lemma 6 we obtain for all $t\in[0,T]$ that

$$\mathbb{E}\left(\|\mu_t^K-\mu_t\|_{\mathrm{BL}^*}\right) \le C_T\Big(R_{d,q}(K)^{\frac12}+I_4(K)\Big),$$

since $I_2(K)\le I_4(K)$, concluding thus the proof. $\square$

We end this section proving the conditional propagation of chaos property stated in Corollary 8.

Proof of Corollary 8. Let $\Psi_{d,q}(K)$ denote the function of K appearing on the right hand side of the bound in Theorem 3. By exchangeability of $\big((X_t^1,Y_t^1),\dots,\big(X_t^{N_t^K},Y_t^{N_t^K}\big)\big)$ conditionally on N_t^K, for all $t\ge 0$ we get

$$\mathbb{E}\Big(\frac{N_t^K}{K}\|X_t^1-Y_t^1\|^2\Big) = \mathbb{E}\Big(\frac{1}{K}\sum_{n=1}^{N_t^K}\|X_t^n-Y_t^n\|^2\Big) \le C_t\Psi_{d,q}^2(K), \tag{5.7}$$

thanks to (5.6). By Proposition 16, we have $\mathcal{L}\big(Y_t^1,\dots,Y_t^j \mid N_t^K\big) = \bar\mu_t^{\otimes j}$ on the event $\{j \le N_t^K\}$. Now, letting $c_t := \langle \mu_t, 1\rangle \in (0,\infty)$ denote the limit in law of N_t^K/K, and using the second inequality of Lemma 4 in the third bound below we get, for all $\varepsilon > 0$, that

$$\begin{aligned}
&\mathbb{P}\Big(\Big\|\mathcal{L}\Big(X_t^1,\dots,X_t^{j\wedge N_t^K}\;\Big|\; N_t^K\Big) - \bar\mu_t^{\otimes j}\Big\|_{\mathrm{BL}^*} > \varepsilon,\, N_t^K \ge j\Big)\\
&\qquad\le \mathbb{P}\bigg(\frac{N_t^K}{K}\Big\|\mathcal{L}\Big(X_t^1,\dots,X_t^j \;\Big|\; N_t^K\Big) - \bar\mu_t^{\otimes j}\Big\|_{\mathrm{BL}^*}\bigg(\frac{N_t^K}{K}\bigg)^{-1} > \frac{\varepsilon c_t}{2}\frac{2}{c_t},\, N_t^K \ge j\bigg)\\
&\qquad\le \mathbb{P}\bigg(\frac{N_t^K}{K}\Big\|\mathcal{L}\Big(X_t^1,\dots,X_t^j \;\Big|\; N_t^K\Big) - \bar\mu_t^{\otimes j}\Big\|_{\mathrm{BL}^*} > \frac{\varepsilon c_t}{2},\, N_t^K \ge j\bigg)\\
&\qquad\quad + \mathbb{P}\bigg(\frac{N_t^K}{K} < \frac{c_t}{2}\bigg)\\
&\qquad\le \frac{2}{\varepsilon c_t}\mathbb{E}\bigg(\frac{N_t^K}{K}\mathbb{E}\bigg(\sum_{n=1}^{j}\|X_t^n - Y_t^n\| \;\bigg|\; N_t^K\bigg)\mathbf{1}_{\{N_t^K\ge j\}}\bigg) + \mathbb{P}\bigg(\frac{N_t^K}{K} < \frac{c_t}{2}\bigg)\\
&\qquad\le \frac{2j}{\varepsilon c_t}\mathbb{E}\Big(\frac{N_t^K}{K}\|X_t^1 - Y_t^1\|\Big) + \mathbb{P}\bigg(\frac{N_t^K}{K} < \frac{c_t}{2}\bigg)\\
&\qquad\le \frac{2j}{\varepsilon c_t}C_t'\Psi_{d,q}^2(K) + \mathbb{P}\bigg(\frac{N_t^K}{K} < \frac{c_t}{2}\bigg),
\end{aligned}$$

using also the Cauchy-Schwarz inequality, the estimate (5.7) and the fact that $\mathbb{E}(N_t^K/K)^{1/2} < \infty$ in the last inequality. Since $N_t^K/K \to c_t$ in law, the last term goes to 0 when $K \to \infty$. The convergences $\mathbb{P}(N_t^K \ge j) \to 1$ and $\Psi_{d,q}(K) \to 0$ as $K \to \infty$ then yield

$$\mathbb{P}\Big(\Big\|\mathcal{L}\Big(X_t^1,\dots,X_t^j \;\Big|\; N_t^K\Big) - \bar\mu_t^{\otimes j}\Big\|_{\mathrm{BL}^*} > \varepsilon \;\Big|\; N_t^K \ge j\Big) \longrightarrow 0,$$

as $K \to \infty$, as required. □

6 Extensions

We end with some remarks regarding possible extensions of our approach, and the technical issues that must be solved in order to establish similar results in some related, more general settings.

Remark 27. If instead of (H.1) it is assumed that the initial data μ_0^K satisfies the condition in Lemma 9 b), the arguments and construction leading to the proof of Theorem 3 must be modified, along the following lines:

- In condition (C.1), $\nu_0^K = \mu_0^K$ is not enforced, but $K\langle \nu_t^K, 1\rangle = K\langle \mu_t^K, 1\rangle = N_t^K$ is kept.
- In the construction of the coupling using algorithm (A), the random variables $(Y^k)_{k\ge1}$ are chosen as before while, for any K and N, the random vectors $(X_0^1,\dots,X_0^N)$ are chosen on the event $\{N_0^K = N\}$, suitably coupled with $(Y_0^1,\dots,Y_0^N)$. This results in an extra term of the form $\mathbb{E}(\langle \eta_0^K, d_2\rangle)$ on the r.h.s. of the bounds in the statement and proof of Lemma 25 which in turn translates into an additional term $C_T\mathbb{E}(\langle \eta_0^K, d_2\rangle)^{1/2}$ on the r.h.s. of the bound in Theorem 3.
- In order to minimize the value of this additional term, the coupling of the variables $(X_0^1,\dots,X_0^N)$ and $(Y_0^1,\dots,Y_0^N)$ must be chosen on each event $\{N_0^K = N\}$ so as to realize the squared 2-Wasserstein distance between the laws of $(X_0^1,\dots,X_0^N)$ and $\bar\mu_0^{\otimes N}$ in $(\mathbb{R}^d)^N$. Denoting

$$\widetilde{W}_2^2(\mathcal{L}(X_0^1,\dots,X_0^N), \bar\mu_0^{\otimes N}) = \frac{1}{N}W_2^2(\mathcal{L}(X_0^1,\dots,X_0^N), \bar\mu_0^{\otimes N}),$$

the normalized squared 2-Wasserstein distance, the additional term $\mathbb{E}(\langle \eta_0^K, d_2\rangle)^{1/2}$ then writes

$$\mathbb{E}\bigg(\frac{N_0^K}{K}\widetilde{W}_2^2\Big(\mathcal{L}\big(X_0^1,\dots,X_0^{N_0^K} \mid N_0^K\big),\bar{\mu}_0^{\otimes N_0^K}\Big)\bigg)^{1/2}.$$

The ideas and techniques developed in this work can in principle also be extended to more general systems of interacting branching populations, including the general setting of [12]. Nevertheless, this requires to deal with significant additional technicalities, and we have chosen to focus here on the basic ideas. The following possible generalizations are left for future work:

- The case of populations with spatially or density depending birth or death events, as in the more general setting studied in [12], seems feasible but presents one major additional difficulty, namely that the jump times are correlated with the spatial dynamics. The main consequence of this is that, in any coupling with some auxiliary system of conditionally independent (or less dependent) particles, the jump times cannot be expected to happen simultaneously. However, under the condition of spatial Lipschitz continuity of the reproduction rate and the competition kernel, it should be possible to keep at least some subsystems effectively coupled on finite time intervals, while controlling explicitly the discrepancy between jump times in the two systems, in terms of the distance of the empirical measures of the systems themselves, in such a way that the discrepancies asymptotically vanish as the population size goes to infinity.
- A further desirable generalization regards the case of branching events more general than binary ones. The natural extension of the argument used here would consist in coupling all the offspring of a branching particle in the original system, with a set of equally many independent new particles given birth at the same time in the auxiliary system. However it is not clear how to make compatible the use of optimal transport plans to couple the branching particle and the positions of the new particles in the auxiliary system, with the independence requirement in the auxiliary system. A possible way of coping with this problem could be to make a two-steps coupling construction: first, between the branching particle in the original system and the positions of new particles in the auxiliary system (which would define an exchangeable random vector of particles in any case) and, in a second step, coupling those positions with independent particles with the required law.

Appendix

Proof of Lemma 4. Since $\|\bar{\nu}\|_{\mathrm{BL}^*} = \langle \bar{\nu}, 1\rangle = 1$, we have

$$\begin{aligned}\|\mu-\nu\|_{\mathrm{BL}^*} &= \|\langle \mu,1\rangle\,(\bar{\mu}-\bar{\nu}) + \bar{\nu}\,(\langle\mu,1\rangle - \langle\nu,1\rangle)\|_{\mathrm{BL}^*}\\ &\le \langle\mu,1\rangle\|\bar{\mu}-\bar{\nu}\|_{\mathrm{BL}^*} + \big|\langle\mu,1\rangle-\langle\nu,1\rangle\big|.\end{aligned}$$

Now, for any $\mu,\nu\in\mathcal{P}(\mathbb{R}^d)$, $\|\mu-\nu\|_{\mathrm{BL}^*} = \sup_{\|\varphi\|_{\mathrm{BL}}\le 1}\left|\int_{\mathbb{R}^d\times\mathbb{R}^d}(\varphi(x)-\varphi(y))\,\pi(\mathrm{d}x,\mathrm{d}y)\right|$ for all coupling $\pi\in\mathcal{P}(\mathbb{R}^{2d})$ of μ and ν. Using the fact that $|\varphi(x)-\varphi(y)|\le|x-y|\wedge 2$ when $\|\varphi\|_{\mathrm{BL}}\le 1$ and taking infimum over all $\pi\in\Pi(\mu,\nu)$ we conclude that $\|\mu-\nu\|_{\mathrm{BL}^*}\le \inf_{\pi\in\Pi(\mu,\nu)}\int|x-y|\wedge 2\,\pi(\mathrm{d}x,\mathrm{d}y)\le W_1(\mu,\nu)$. □

Proof of Lemma 6. Write $\alpha = 1/2$ when $d<4$ or $\alpha = 2/d$ when $d>4$. Thanks to

Theorem 5, for some $C_{d,q} > 0$,

$$\begin{aligned}\mathbb{E}\Big(\frac{N}{K}W_2^2(\bar\nu^K,\bar\mu)\Big) &= \mathbb{E}\Big(\frac{N}{K}\mathbb{E}\Big(W_2^2(\bar\nu^K,\bar\mu)\ \Big|\ N\Big)\Big)\\ &\le C_{d,q}M_q^{\frac{2}{q}}(\bar\mu)\,\mathbb{E}\Big(\frac{N}{K}\Big(N^{-\alpha}+N^{-\frac{q-2}{q}}\Big)\Big)\\ &= C_{d,q}M_q^{\frac{2}{q}}(\bar\mu)\Big(K^{-\alpha}\,\mathbb{E}\Big(\Big(\frac{N}{K}\Big)^{1-\alpha}\Big)+K^{-\frac{q-2}{q}}\mathbb{E}\Big(\Big(\frac{N}{K}\Big)^{\frac{2}{q}}\Big)\Big)\\ &\le C_{d,q}M_q^{\frac{2}{q}}(\bar\mu)\Big(K^{-\alpha}\mathbb{E}\Big(\frac{N}{K}\Big)^{1-\alpha}+K^{-\frac{q-2}{q}}\mathbb{E}\Big(\frac{N}{K}\Big)^{\frac{2}{q}}\Big),\end{aligned}$$

using Jensen's inequality in the last line. This implies the result for $d\neq 4$. When $d=4$ we get the bounds

$$\begin{aligned}\mathbb{E}\Big(\frac{N}{K}W_2^2(\bar\nu^K,\bar\mu)\Big) &\le C_{d,q}M_q^{\frac{2}{q}}(\bar\mu)\Big(K^{-\frac12}\,\mathbb{E}\Big(\Big(\frac{N}{K}\Big)^{\frac12}\log(1+N)\Big)+K^{-\frac{q-2}{q}}\mathbb{E}\Big(\frac{N}{K}\Big)^{\frac{2}{q}}\Big)\\ &\le C_{d,q}M_q^{\frac{2}{q}}(\bar\mu)\Big(K^{-\frac12}\,\mathbb{E}\Big(\frac{N}{K}\Big)^{\frac12}\mathbb{E}\Big(\log^2(e+N)\Big)^{\frac12}+K^{-\frac{q-2}{q}}\mathbb{E}\Big(\frac{N}{K}\Big)^{\frac{2}{q}}\Big).\end{aligned}$$

The function $x\in[e,\infty)\mapsto\log^2(x)$ being concave, it can be extended linearly on $(-\infty,e)$ to get a C^1 concave function on $\mathbb{R}$. Jensen's inequality then yields

$$\mathbb{E}\Big(\log^2(e+N)\Big)^{\frac12}\le\log\Big(e+K\mathbb{E}\Big(\frac{N}{K}\Big)\Big)\le 1+\log(1+K)+\log\Big(1\vee\mathbb{E}\Big(\frac{N}{K}\Big)\Big).$$

Using this bound, we finally obtain that

$$\begin{aligned}\mathbb{E}\Big(\frac{N}{K}W_2^2(\bar\nu^K,\bar\mu)\Big) &\le C_{d,q}M_q^{\frac{2}{q}}(\bar\mu)\Big(K^{-\frac12}\,\mathbb{E}\Big(\frac{N}{K}\Big)^{\frac12}+K^{-\frac12}\log(1+K)\,\mathbb{E}\Big(\frac{N}{K}\Big)^{\frac12}\\ &\quad+K^{-\frac12}\,\mathbb{E}\Big(\frac{N}{K}\Big)^{\frac12}\log\Big(1\vee\mathbb{E}\Big(\frac{N}{K}\Big)\Big)+K^{-\frac{q-2}{q}}\mathbb{E}\Big(\frac{N}{K}\Big)^{\frac{2}{q}}\Big),\end{aligned}$$

and the case $d=4$ follows since $K^{-\frac12}\le K^{-\frac12}\log(1+K)$ for $K\in\mathbb{N}\setminus\{0\}$. □

Proof of Lemma 9. Since condition (H.1) assumed in a) is a particular case of the assumptions in b), it is enough to prove b) to get both parts. Denote by $m\in(0,\infty)$ the limit in law of $(\langle\mu_0^K,1\rangle)_{K\in\mathbb{N}\setminus\{0\}}$. Taking $\mu=m\tilde\mu_0$ and $\nu=\mu_0^K$ in Lemma 4, we get

$$\limsup_K\mathbb{P}(\|m\tilde\mu_0-\mu_0^K\|_{\mathrm{BL}^*}\ge\varepsilon)\le\limsup_K\mathbb{P}(\|\tilde\mu_0-\bar\mu_0^K\|_{\mathrm{BL}^*}\ge\varepsilon/(2\langle m\tilde\mu_0,1\rangle)),\tag{A.1}$$

with $\bar\mu_0^K=\frac{1}{N_0^K}\sum_{i=1}^{N_0^K}\delta_{X_0^i}$. On the other hand, for each $\delta>0$ and $M>0$,

$$\begin{aligned}\mathbb{P}(\|\tilde\mu_0-\bar\mu_0^K\|_{\mathrm{BL}^*}\ge\delta) &\le\sum_{N\ge M}\mathbb{E}\left[\mathbb{P}(\|\tilde\mu_0-\bar\mu_0^K\|_{\mathrm{BL}^*}\ge\delta|N_0^K=N)\mathbf{1}_{N_0^K=N}\right]+\mathbb{P}(N_0^K<M)\\ &\le\sup_{N\ge M}\mathbb{P}\left(\left\|\tilde\mu_0-\frac{1}{N}\sum_{i=1}^N\delta_{Y^{i,N}}\right\|_{\mathrm{BL}^*}\ge\delta\right)+\mathbb{P}(\langle\mu_0^K,1\rangle<M/K).\end{aligned}$$

Since $\langle\mu_0^K,1\rangle$ converges weakly to a non null constant, the last term goes to 0 when $K\to\infty$. Now, it is well known that the assumed $\tilde\mu_0$-chaoticity is equivalent to the convergence in distribution of the random probability $\frac1N\sum_{i=1}^N\delta_{Y^{i,N}}$ to $\tilde\mu_0$ as $N\to\infty$. If follows that $\limsup_{K\to\infty}\mathbb{P}(\|\tilde\mu_0-\bar\mu_0^K\|\ge\delta)=0$ which entails the claim in view of (A.1).

c) The r.v. $N_0^K=K\langle\mu_0^K,1\rangle$ is Poisson of parameter $K\langle\mu_0,1\rangle$ and equals in law the sum $\sum_{i=1}^K N^i$ of independent Poisson r.v. $(N^i)_{i=1}^K$ of parameter $\langle\mu_0,1\rangle$. By the law of large numbers, $\langle\mu_0^K,1\rangle=N_0^K/K$ converges in law to the constant $\langle\mu_0,1\rangle$. It is immediate from basic properties of Poisson point measures that the N_0^K atoms of μ_0^K are i.i.d. of law $\bar\mu_0$ given $\langle\mu_0^K,1\rangle$. Last, N_0^K being Poisson of parameter $K\langle\mu_0,1\rangle$, we have $I_4^4(K)=K^{-3}\left(\langle\mu_0,1\rangle+3K\langle\mu_0,1\rangle^2\right)\le CK^{-2}$. □

References

[1] Vincent Bansaye and Sylvie Méléard, *Stochastic models for structured populations*, Mathematical Biosciences Institute Lecture Series. Stochastics in Biological Systems, vol. 1, Springer, Cham; MBI Mathematical Biosciences Institute, Ohio State University, Columbus, OH, 2015, Scaling limits and long time behavior. MR3380810

[2] José A. Carrillo, Yanghong Huang, and Markus Schmidtchen, *Zoology of a nonlocal cross-diffusion model for two species*, SIAM J. Appl. Math. **78** (2018), no. 2, 1078–1104. MR3783102

[3] José Antonio Carrillo, Young-Pil Choi, and Maxime Hauray, *The derivation of swarming models: mean-field limit and Wasserstein distances*, Collective dynamics from bacteria to crowds, CISM Courses and Lect., vol. 553, Springer, Vienna, 2014, pp. 1–46. MR3331178

[4] Li Chen, Esther S. Daus, Alexandra Holzinger, and Ansgar Jüngel, *Rigorous derivation of population cross-diffusion systems from moderately interacting particle systems*, J. Nonlinear Sci. **31** (2021), no. 6, Paper No. 94, 38. MR4319922

[5] Li Chen, Esther S. Daus, and Ansgar Jüngel, *Rigorous mean-field limit and cross-diffusion*, Z. Angew. Math. Phys. **70** (2019), no. 4, Paper No. 122, 21. MR3989190

[6] Xiuqing Chen, Esther S. Daus, and Ansgar Jüngel, *Global existence analysis of cross-diffusion population systems for multiple species*, Arch. Ration. Mech. Anal. **227** (2018), no. 2, 715–747. MR3740386

[7] Roberto Cortez and Joaquin Fontbona, *Quantitative propagation of chaos for generalized Kac particle systems*, Ann. Appl. Probab. **26** (2016), no. 2, 892–916. MR3476628

[8] Roberto Cortez and Joaquin Fontbona, *Quantitative uniform propagation of chaos for Maxwell molecules*, Comm. Math. Phys. **357** (2018), no. 3, 913–941. MR3769742

[9] Donald A. Dawson, *Measure-valued Markov processes*, École d'Été de Probabilités de Saint-Flour XXI—1991, Lecture Notes in Math., vol. 1541, Springer, Berlin, 1993, pp. 1–260. MR1242575

[10] M. Di Francesco, A. Esposito, and S. Fagioli, *Nonlinear degenerate cross-diffusion systems with nonlocal interaction*, Nonlinear Anal. **169** (2018), 94–117. MR3761096

[11] Marco Di Francesco and Simone Fagioli, *Measure solutions for non-local interaction PDEs with two species*, Nonlinearity **26** (2013), no. 10, 2777–2808. MR3105514

[12] Joaquin Fontbona and Sylvie Méléard, *Non local Lotka-Volterra system with cross-diffusion in an heterogeneous medium*, J. Math. Biol. **70** (2015), no. 4, 829–854. MR3306618

[13] Nicolas Fournier and Arnaud Guillin, *On the rate of convergence in Wasserstein distance of the empirical measure*, Probab. Theory Related Fields **162** (2015), no. 3-4, 707–738. MR3383341

[14] Nicolas Fournier and Sylvie Méléard, *A microscopic probabilistic description of a locally regulated population and macroscopic approximations*, Ann. Appl. Probab. **14** (2004), no. 4, 1880–1919. MR2099656

[15] Sylvie Méléard, *Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models*, Probabilistic models for nonlinear partial differential equations (Montecatini Terme, 1995), Lecture Notes in Math., vol. 1627, Springer, Berlin, 1996, pp. 42–95. MR1431299

[16] Nanako Shigesada, Kohkichi Kawasaki, and Ei Teramoto, *Spatial segregation of interacting species*, J. Theoret. Biol. **79** (1979), no. 1, 83–99. MR540951

[17] Alain-Sol Sznitman, *Topics in propagation of chaos*, École d'Été de Probabilités de Saint-Flour XIX—1989, Lecture Notes in Math., vol. 1464, Springer, Berlin, 1991, pp. 165–251. MR1108185

[18] Cédric Villani, *Optimal transport*, Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 338, Springer-Verlag, Berlin, 2009, Old and new. MR2459454

Acknowledgments. We thank an anonymous referee for helpful comments and suggestions that allowed us to improve the presentation of the paper, and Roberto Cortez for carefully reading the final version of the manuscript.