Latent Voter Model on Locally Tree-Like Random Graphs

Ran Huo and Rick Durrett *
Dept. of Math, Duke U.,
P.O. Box 90320, Durham, NC 27708-0320

March 11, 2017

Abstract

In the latent voter model, which models the spread of a technology through a social network, individuals who have just changed their choice have a latent period, which is exponential with rate λ, during which they will not buy a new device. We study site and edge versions of this model on random graphs generated by a configuration model in which the degrees $d(x)$ have $3 \le d(x) \le M$. We show that if the number of vertices $n \to \infty$ and $\log n \ll \lambda_n \ll n$ then the fraction of 1's at time $\lambda_n t$ converges to the solution of $du/dt = c_p u(1-u)(1-2u)$. Using this we show the latent voter model has a quasi-stationary state in which each opinion has probability $\approx 1/2$ and, with high probability, persists in this state for a time that is $\ge n^m$ for any $m < \infty$. Thus, even a very small latent period drastically changes the behavior of the voter model, which has a one parameter family of stationary distributions and reaches fixation in time $O(n)$.

1 Introduction

In this paper we will study the latent voter model introduced in 2009 by Lambiotte, Saramaki, and Blondel [13]. In this model each individual owns one of two types of technology, say an iPad or a Microsoft Surface tablet. In the voter model on the d-dimensional lattice, individuals at times of a rate one Poisson process pick a neighbor at random and imitate their opinion. However, in the current interpretation of that model, it is unlikely that someone who has recently bought a new tablet computer will replace it, so we introduce latent states 1^* and 2^* in which individuals will not change their opinion. If an individual is in state 1 or 2 we call them active. Letting f_i be the fraction of neighbors in state i or i^*, the dynamics can be formulated as follows

$$\begin{array}{ll} 1 \to 2^* \text{ at rate } f_2 & \qquad 1^* \to 1 \text{ at rate } \lambda \\ 2 \to 1^* \text{ at rate } f_1 & \qquad 2^* \to 2 \text{ at rate } \lambda \end{array}$$

*RD is partially supported by NSF grant DMS 1505215 from the probability program.

In [13] the authors showed that if individuals in the population interact equally with all the others then the system converges to a limit in which both technologies have frequency close to 1/2. Here, we will study the system with large λ, since in this case it is a voter model perturbation in the sense of Cox, Durrett, and Perkins [5]. To explain this, we will construct the system using a graphical representation. Suppose first that the system takes place on $\mathbb{Z}^d$ and that $d \geq 3$. For each $x \in \mathbb{Z}^d$ and nearest neighbor y, we have independent Poisson processes $T_n^{x,y}$, $n \geq 1$. At each time $t = T_n^{x,y}$ we draw an arrow from $(y,t) \to (x,t)$ to indicate that if the individual at x is active at time t then they will imitate the opinion at y.

To implement the other part of the mechanism, we introduce for each site x, a Poisson process T_n^x, $n \geq 1$ of "wake-up dots" that return the voter to the active state.

- If there is only one voter arrow between two wake up dots, the result is an ordinary voter event.
- If between two wake up dots there are voter arrows to x from two different neighbors, an event of probability $O(\lambda^{-2})$, then x will change its opinion if and only at least one of the two neighbors has a different opinion. To check this, we note that if the first arrow causes a change then the second one is ignored, while if the first arrow comes from a site with the same opinion as the one at x then there will be a change if and only if the second site has an opinion different from the one at x.
- If t is fixed then at a given site there are $O(\lambda)$ wake-up dots by time t. Thus if we want to see the influence of intervals with two voter arrows then we want to run time at rate λ. The probability of k voter arrows between two wake-up dots is $(1+\lambda)^{-k}$, so in the limit the probability of three of more voter events between two wake-up dots goes to 0 as $\lambda \to \infty$.

If we let $\lambda = \varepsilon^{-2}$ and let $n_k(x)$ be the number of neighbors in state k or k^* then the rate of flips from 1 to 2 in the latent voter model when the configuration is ξ is:

$$\varepsilon^{-2} c^v_{1,2}(x,\xi) + h_{1,2}(x,\xi) \quad \text{where} \quad c^v_{1,2}(x,\xi) = 1_{\{\xi(x)=1\}} \frac{n_2(x)}{2d}$$

If we let $y_1, \dots y_{2d}$ be an enumeration of the nearest neighbors of x, the perturbation is

$$h_{1,2}(x,\xi) = 1_{\{\xi_t(x)=1\}} \frac{2}{(2d)^2} \sum_{1 \leq k < \ell \leq 2d} 1_{\{\xi(y_k) \text{ or } \xi(y_\ell) \in \{2,2^*\}\}}$$

Similar formulas hold when the roles of 1 and 2 are interchanged. $h_{1^*,j} = h_{2^*,j} \equiv 0$.

If we scale space by ε then Theorem 1.2 of [5] shows that under mild assumptions on the perturbation, the density of 1's in the rescaled particle system converges to the solution of the limiting PDE:

$$\frac{\partial u}{\partial t} = \frac{1}{2}\Delta u + \phi(u) \quad \text{with} \quad \phi(u) = \langle h_{2,1}(0,\xi) - h_{1,2}(0,\xi) \rangle_u \tag{1}$$

and $\langle \cdot \rangle_u$ denotes the expected value with respect to the voter model with density u.

Intuitively, (1) holds because of a separation of time scales. The rapid voting means that the configuration near x looks like the voter model equilibrium with density $u(t,x)$. Later in the paper will show, see (11), that in the case of the latent voter model

$$\phi(u) = c_d u(1-u)(1-2u).$$

If we consider the latent voter model on a torus with n sites and let $\lambda_n \to \infty$ then the system can be analyzed using ideas from a recent paper of Cox and Durrett [4]. Define the density of 1's at time t by

$$U^n(t) = \frac{1}{n}\sum_x 1_{\{\xi_{\lambda t}(x)=1\}} \tag{2}$$

Theorem 1. *Suppose $n^{2/d} \ll \lambda_n \ll n$. If $U^n(0) \to u_0$ then $U^n(t)$ converges uniformly on compact sets to $u(t)$ the solution of*

$$\frac{du}{dt} = c_d u(1-u)(1-2u) \qquad u(0) = u_0$$

Remark 1. Note that the only thing we assume about the initial state is that the density $U^n(0) \to u_0$. Fast voting will turn the initial condition into a voter model equilibrium in a time that is $o(\lambda_n)$. If we consider the voter model without a latent period then the limiting differential equation is $du/dt = 0$. The last conclusion for the voter model is a very simple special case of the results in [4].

1.1 Random graphs

We will explain the intuition behind Theorem 1 after we state our new result that replaces the torus by a random graph G_n generated by the *configuration model*. For the rest of the paper we will only consider the latent voter model on G_n. In the configuration model vertices have degree k with probability p_k. To create that graph we assign i.i.d. degrees d_i to the vertices and condition the sum $d_1 + \cdots + d_n$ to be even, which is a necessary condition for the values to be the degrees of a graph. We attach d_i half-edges to vertex i and then pair the half-edges at random. We will assume that

(A0) the graph G_n has no self-loops or parallel edges.

If $\sum_k k^2 p_k < \infty$ then the probability of (A0) is bounded away from 0 as $n \to \infty$. See Theorem 3.1.2 of [8]. The reader can consult Chapter 3 of that reference for more on the configuration model.

It seems likely that the results we prove here are true under the assumption that the degree distribution has finite second moment, but the presence of vertices of large degrees causes a number of technical problems. To avoid these we will assume:

(A1) $p_m = 0$ for $m > M$, i.e., the degree distribution is bounded.

In addition, we want a graph that is connected and has random walks with good mixing times, so we will also suppose:

(A2) $p_k = 0$ for $k \le 2$.

The relevance of (A2) for mixing times will be explained in Section 2. Assumptions (A0), (A1) and (A2) will be in force throughout the paper.

On graphs that are not regular there are two versions of the voter model.

(i) The *site version* in which sites change their opinions at rate 1, and imitate a neighbor chosen at random,

$$c^s_{i,j}(x,\xi) = 1_{\{\xi(x)=i\}} \frac{n_j(x)}{d(x)}$$

where $n_j(x)$ is the number of neighbors of x in state j, and $d(x)$ is the degree of x.

(ii) The *edge version* in which each neighbor that is different from x causes its opinion to change at rate 1,

$$c^e_{i,j}(x,\xi) = 1_{\{\xi(x)=i\}} n_j(x).$$

The site version is perhaps the "obvious" generalization of the voter model on $\mathbb{Z}^d$, e.g., it is a special case of the general formulation used in Liggett [15]: x imitates y with probability $p(x,y)$, where p is a transition probability. However, the edge version has two special properties. First, in the words of [23] "magnetization is conserved," i.e., the number of 1's is a martingale. Second, if we consider the biased version in which after an edge (x,y) is picked a 1 at x always imitate a 2 at y but a 2 at x imitates a 1 at y with probability $\rho < 1$ then the probability a single 2 takes over a system that is otherwise all 1 is the same as the probability a simple random walk that jumps up with probability $1/(1+\rho)$ and down with probability $\rho/(1+\rho)$ never hits 0. This observation is due to Maruyama in 1970 [17], but has recently been rediscovered by [14], who call this version of the voter model "isothermal".

From our discussion of the graphical representation for latent voter model on $\mathbb{Z}^d$ it should be clear that the latent voter model on G_n is a voter model perturbation. If we let $y_1, \dots y_{d(x)}$ be an enumeration of the neighbors of x, then in the site version

$$h^s_{1,2}(x,\xi) = 1_{\{\xi_t(x)=1\}} \frac{2}{(d(x))^2} \sum_{1 \le k < \ell \le d(x)} 1_{\{\xi(y_k) \text{ or } \xi(y_\ell) \in \{2,2^*\}\}}$$

while in the edge version

$$h^e_{1,2}(x,\xi) = 1_{\{\xi_t(x)=1\}} \cdot 2 \sum_{1 \le k < \ell \le d(x)} 1_{\{\xi(y_k) \text{ or } \xi(y_\ell) = \in \{2,2^*\}\}}$$

As before interchanging the roles of 1 and 2 we can define $h^s_{2,1}(x,\xi)$ and $h^e_{2,1}(x,\xi)$ while $h^s_{i,j}(x,\xi) = h^e_{i,j}(x,\xi) = 0$ when $i = 1^*$ or 2^*.

The last detail is to define the density $U^n(t)$. To do this we note that a random walk that jumps to a neighbor chosen at random has stationary distribution $\pi(x) = d(x)/D$, where $D = \sum_y d(y)$, while a random walk that jumps to each neighbor at rate 1 has stationary distribution $\pi(x) = 1/n$. To treat the two cases in one definition we let

$$U^n(t) = \sum_x \pi(x) 1_{\{\xi_{\lambda t}(x)=1\}} \tag{3}$$

Theorem 2. *Suppose that* $\log n \ll \lambda_n \ll n$. *If* $U^n(0) \to u_0$ *then* $U^n(t)$ *converges in probability and uniformly on compact sets to* $u(t)$*, the solution of*

$$\frac{du}{dt} = c_p u(1-u)(1-2u) \qquad u(0) = u_0. \tag{4}$$

where the value of c_p *depends on the degree distribution and the version of the voter model.*

Remark 2. Again in the voter model without a latent period the limit is $du/dt = 0$. That result can be easily proved using the arguments for Theorem 2.

1.2 Duality

To explain why Theorems 1 and 2 are true, we will introduce a dual process that is the key to the analysis. The dual process was first introduced more than 20 years ago by Durrett and Neuhauser [11], and is the key to work of Cox, Durrett, and Perkins [5]. To do this, we construct the process using a graphical representation that generalizes the one introduced for $\mathbb{Z}^d$. For each $x \in \mathbb{Z}^d$ and neighbor y, we have independent Poisson processes $T_n^{x,y}$, $n \geq 1$. At each time $t = T_n^{x,y}$ we draw an arrow from $(y,t) \to (x,t)$ to indicate that if the individual at x is active at time t then they will imitate the opinion at y. In the edge case all these processes have rate 1. In the site case $T_n^{x,y}$, $n \geq 0$ has rate $1/d(x)$. To implement the other part of the mechanism, we have for each site x, a rate λ Poisson process T_n^x, $n \geq 1$ of "wake-up dots" that return the voter to the active state.

To compute the state of x at time t we start with a particle at x at time t. To be precise $\zeta_0^{x,t} = \{x\}$. As we work backwards in time the particle does not move until the first time s there is an arrow $(y, t-s) \to (x, t-s)$.

- If this is the only voter arrow between the two adjacent wake-up dots then the particle jumps to y.
- If in the interval between the two adjacent wake-up dots there are arrows from k distinct y_i then the state changes to $\{x, y_1, \ldots y_k\}$ since we need to know the current state of all these points to know what change should occur in the process. In the limit as $\lambda \to \infty$ we will only see branchings that add two y_i. We include the case $k > 2$ to have the dual process well-defined.
- We do not need to know the order of the arrows because x will change if at least one of the y_i has a different opinion. When λ is small some of the y_i might change their state during the interval between the two wake-up dots but this possibility has probability zero in the limit.
- To complete the definition of the dual, we declare that if a branching event adds a point already in the set, or if a particle jumps onto an occupied site then the two coalesce to one.

The dual process can be used to compute the state of x at time t. The first step is to work backwards in time to find $\zeta_t^{x,t}$ the set of sites at time 0 that can influence the state of x at time t. We note the states of the sites at time 0 and then work up the graphical representation to determine what changes should occur at the branching points in the dual.

To prove Theorem 1, Cox and Durrett [4] show that after a branching event any coalescence between the particle that branched and the two newly created particles will happen quickly, in time $O(1)$ or these particles will need time $O(n)$ to coalesce. (here we are using the original time scale.) Let $L = n^{1/d}$ be the side length of the torus. When $\lambda_n \gg n^{2/d}$ the particles will come to equilibrium on the torus before the next branching occurs in the dual, so we can forget about the relative location of the particles and we end up with an ODE limit. On the random graph, our assumption that all vertices have degree ≥ 3 implies that the mixing time for random walks on these graphs is $O(\log n)$. Thus when $\lambda_n \gg \log n$, we have the situation that after a branching event there may be some coalescence in the dual at times $O(1)$ but then the existing particles will come to equilibrium on the graph before the next branching occurs in the dual. In both cases $\lambda_n \ll n$ is needed for the perturbation to have a nontrivial effect. Otherwise a collection of k random walks might all coalesce before the first branching arrow.

Remark 3. There is no reason for having vertices of degree 0 in our graph. If $p_2 > 0$ and we look at the dynamics on the giant component then Theorem 2 will hold if $\log^2 n \ll \lambda_n \ll n$. The increase in the lower bound is needed to compensate for the fact that the mixing time for random walks on the graph is $O(\log^2 n)$. See e.g., Section 6.7 in [8]. Allowing $p_1 > 0$ should not change the behavior but, for simplicity, we derive our results under the assumption $d(x) \geq 3$.

1.3 Long time survival

The latent voter model has two absorbing states $\equiv 1$ and $\equiv 2$. On a finite graph the latent voter model is a finite state Markov chain, so we know it will eventually reach one of them. However by analogy with the contact process on the torus [18] and on power-law random graphs, [19], this result should hold for times up to $\exp(\gamma n)$ for some $\gamma > 0$. Unfortunately we are only able to prove survival for any power of n. This failure is to due to our estimate of $P(\Omega_1^c)$ which appears on the right-hand side of Theorem 4.

Theorem 3. *Suppose that* $\log n \ll \lambda_n \ll n$. *Let* $\varepsilon > 0$ *and* $k < \infty$. *If* $U^n(0) \to u_0 \in (0,1)$ *there is a* T_0 *that depends on the initial density so that for any* $m < \infty$ *if* n *is large then with high probability*

$$|U^n(t) - 1/2| \leq 5\varepsilon \qquad \textit{for all } t \in [T_0, n^m].$$

Remark 4. Here and in what follows "with high probability" means with probability $\to 1$ as $n \to \infty$. Cox and Greven [6] have shown that for the nearest neighbor voter model on the torus in $d \geq 3$ that if we let θ_t be the fraction of sites in state 1 at time Nt then the configuration at time nt looks like the voter model equilibrium with density θ_t and θ_t changes according to the Wright-Fisher diffusion

$$d\theta_t = \sqrt{\beta_d \cdot 2\theta_t(1-\theta_t)}\, dB_t$$

with β_d the probability that two random walks starting from adjacent sites never hit.

To prove Theorem 3, we use Theorem 4.2 of Darling and Norris [7]. To state their result we need to introduce some notation. To make it easier compare with their paper we use their notation even though in some cases it conflicts with ours. Let ξ_t be a continuous time Markov chain with countable state space S and jump rates $q(\xi,\xi')$. In our case ξ_t will be the state of the voter model on the random graph. For their coordinate function $x : S \to \mathbb{R}^d$ we will take $d = 1$ and

$$x(\xi) = \sum_{x\in G_n} \pi(x) 1_{\{\xi(x)=1\}}.$$

We are interested in proving an ODE limit for $X_t = x(\xi_{\lambda t})$. Here and in what follows we drop the subscript n. To compare with the paper note that our ξ_t is their X_t and our X_t is their $\mathbf{X}_t$.

If we set $U = [0, 1]$ in [7] then we always have $x(\xi_t) \in U$ so their condition (2) is not needed. For each $\xi \in S$ we define the drift vector

$$\beta(\xi) = \sum_{\xi'\neq\xi} (x(\xi') - x(\xi))q(\xi, \xi')$$

We let b be the drift of the proposed deterministic limit x_t:

$$x_t = x_0 + \int_0^t b(x_s)\, ds.$$

In our case $b(y) = cy(1 - y)(1 - 2y)$. To measure the size of the jumps we let $\sigma_\theta(y) = e^{\theta|y|} - 1 - \theta|y|$ and let

$$\phi(\xi,\theta) = \sum_{\xi'\neq\xi} \sigma_\theta(x(\xi') - x(\xi))q(\xi, \xi').$$

Consider the events $\Omega_0 = \{|X_0 - x_0| \le \eta\}$,

$$\Omega_1 = \left\{ \int_0^t |\beta(\xi_{\lambda s}) - b(X_s)|\, ds \le \eta \right\},$$
$$\text{and } \Omega_2 = \left\{ \int_0^t \phi(\xi_s, \theta)\, ds \le \theta^2 At/2 \right\}.$$

The parameters in these events are coupled by the following relationships. If we let K be the Lipschitz constant of the drift b then $\eta = \varepsilon e^{-Kt_0/3}$ and $\theta = \eta/(At)$ where $A > 0$. We have changed their δ to η because we use δ in a number of our arguments in Section 3.

Theorem 4. *Under the conditions above, for each fixed t*

$$P\left(\sup_{s\le t_0} |X_s - x_s| > \varepsilon\right) \le 2e^{-\eta^2/(2At_0)} + P(\Omega_0^c \cup \Omega_1^c \cup \Omega_2^c)$$

In our application t_0 will be fixed and K is independent of n so η does not depend on n. To make the first term vanish in the limit we will take $A = n^{-1/2}$. To bound the probabilities on the right-hand side, we note

- We have jumps that change the density by $1/n$ at times of a Poisson processes at total rate $\leq M\lambda n$. If $\theta|y|$ is small $\sigma_\theta(\pm 1/n) \sim \theta^2/2n^2$. So if n is large enough so that $M\lambda/n \leq n^{-1/2}t_0/2$ then a standard large deviations estimate for the $P(\Omega_2^c) \leq \exp(-c\lambda n)$.

- Our assumption that $U^n(0) \to u_0$ implies that $\Omega_0^c = \emptyset$ for large n.

- The hard work comes in estimating $P(\Omega_1^c)$, i.e., estimating the difference in the drift in the particle system from what we compute on the basis of the current density. We do this in Section 3.3-3.4 by computing the expected value of the mth moment of the difference $|\beta(\xi_s) - b(X_s)|$ so we end up with estimates that for a fixed time are $\leq n^{-m}$.

Once these three steps are done, the remainder of the proof of Theorem 3 given in Section 3.5 is routine.By subdividing the interval into small pieces we can use the single time estimates to control the supremum and hence the integral but only over a bounded time interval. However this is enough since it allows us to show that when the density wanders more than 4ε away from $1/2$, we can return it to within 2ε with probability n^{-m}, and in addition never have the difference exceed 5ε.

Theorem 2 is proved in Section 2 and Theorem 3 in Section 3. These results hold for other voter model perturbations such as the evolutionary games considered in [4]. However, the main obstacle to proving a general result is to find a formulation that works well on graphs with variable degrees. The arguments in the first proof closely parallel arguments in [4] but now use estimates for random walks on random graphs.

The keys to the second proof are results concerning the behavior of coalescing random walks (CRWs). There have been a number of studies of the time it takes for CRWs starting from every site of a random graph to coalesce to 1. Cooper et al [2, 3] and Oliveira [20] considered coalescing random walks with one particle at each site and obtained results on the time needed for all particles to coalesce to 1. In [21] sufficient conditions were given for the number of particles in the coalescing random walk to converge to Kingman's coalescent. However, here we need estimates on the number of coalescences that can occur by time $C \log n$. This is done in Section 3.2.

2 Proof of Theorem 2

2.1 Mixing times for random walks

Bounds for the mixing times come from studying the conductance

$$Q(x, y) = \pi(x)q(x, y)$$

where π is the stationary distribution and $q(x, y)$ is the rate of jumping from x to y. In the site version $q(x, y) = 1/d(x)$ while $\pi(x) = d(x)/D$ when y is a neighbor of x, $y \sim x$, so $Q(x, y) = 1/D$ when $y \sim x$. In the edge version, $q(x, y) = 1$ if $y \sim x$, while $\pi(x) = 1/n$ where n is the number of vertices, so $Q(x, y) = 1/n$ when $y \sim x$. When degrees are bounded, the two conductances are the same up to a constant.

Define the isoperimetric constant by

$$h = \min_{\pi(S) \le 1/2} \frac{Q(S, S^c)}{\pi(S)}$$

where $\pi(S) = \sum_{x \in S} \pi(x)$ and $Q(S, S^c) = \sum_{x \in S, y \in S^c} Q(x, y)$. Cheeger's inequality, see e.g. Theorem 6.2.1. in [8] implies that the spectral gap of Q, $\beta = 1 - \lambda_1$ has

$$\frac{h^2}{2} \le \beta \le 2h \tag{5}$$

Using Theorem 6.1.2 in [8] we see that if $p_t(x, y)$ is the transition probability associated with Q

$$\Delta(t) \equiv \max_{x,y} \left| \frac{p_t(x, y)}{\pi(y)} - 1 \right| \le \frac{e^{-\beta t}}{\pi_{min}} \tag{6}$$

where $\pi_{min} = \min \pi(x)$.

Gkantsis, Mihail, and Saberi [12] have shown, see Theorem 6.3.2. in [8]:

Theorem 5. *Consider a random graph in which the minimum degree is ≥ 3. There is a constant α_0 so that with high probability $h \ge \alpha_0$.*

Combining the last result with (5), (6), and the fact that $\pi_{min} \ge 1/(C_0 n)$ for large n, we see that

$$\Delta(t) \le C_0 n e^{-\gamma t} \qquad \text{where } \gamma = \alpha_0^2/2.$$

If we let $C_1 = (6/\alpha_0^2)$ then n large we have for $t \ge C_1 \log n$

$$\Delta(t) \le 1/n \tag{7}$$

2.2 Our random graph is (almost) locally a tree

Recall that to construct our random graph we let $d_1, d_2, \dots d_n$ be i.i.d. from the degree distribution conditioned on $d_1 + \cdots + d_n$ to be even and then we pair the half-edges at random. Given a vertex x with degree $d(x)$, we let $y_1(x) \dots y_{d(x)}(x)$ be its neighbors. To grow the graph we let $V_0 = \{x\}$. On the first step we draw edges from x to $y_1(x) \dots y_{d(x)}(x)$ and let $V_1 = \{y_1(x), \dots, y_{d(x)}(x)\}$ which we consider to be an ordered list. If V_t has been constructed we let x_t be the first element of V_t and draw edges from x_t to $y_1(x_t) \dots y_{d(x_t)}(x_t)$. To create V_{t+1} we remove x_t and add the members of $y_1(x_t) \dots y_{d(x_t)}(x_t)$ not already in V_t.

We stop when we have determined the neighbors of all vertices at distance $< (1/5) \log_M n$ from x. A simple calculation using branching processes shows that the total number of neighbors within that distance of x is $\le n^{1/5} \log n$ for large n. The $\log n$ takes care of the limiting random variable. Thus in the construction we will generate $\le M n^{1/5} \log n$ connections. We say that a collision occurs at time t if we connect to a vertex already in V_t. The probability of a collision on single connection is $\le M n^{-4/5} \log n$. The expected number of collisions involving the first $n^{1/5} \log n$ sites is $\le C M n^{-3/5} \log^2 n$, so for most starting points (but not all) the graph will be a tree. To get a conclusion that applies to all starting points we note that the probability of two collisions in the construction starting from one site is

$$\le \binom{C M n^{1/5} \log n}{2} (n^{-4/5} \log^2 n)^2 = O(n^{-6/5} \log^6 n)$$

As we build up the graph we first find all of the neighbors of vertices at distance 1 from x then distance 2, etc. Thus when a collision occurs it will connect a vertex at distance k with one at distance k or to one at distance $k+1$ that already has a neighbor at distance k. As we will explain after the proof of the next lemma, this makes very little difference.

2.3 Results for hitting times

Lemma 2.1. *Once two particles are at distance r_n with $1 \le r_n \le (1/25)\log n$ then with probability $\ge 1 - 2^{1-r_n}$, they will reach a distance $5r_n$ before hitting each other.*

Proof. For the proof we will pretend that the graph is exactly a tree up to distance $5r_n$. We return to this issue in a remark after the proof. Let Z_t be the distance between these two particles and let T_m be the first time the distance is m. Note that on each jump, with probability $p \ge 2/3$, the particles get 1 step further apart, while with probability $\le 1/3$, the particles get one step closer. This implies that $\phi(z) = (1/2)^z$ is a supermartingale, so

$$\phi(r_n) \ge P_{r_n}(T_0 < T_{5r_n})\phi(0) + (1 - P_{r_n}(T_0 < T_{5r_n}))\phi(5r_n).$$

Rearranging gives

$$P_{r_n}(T_0 < T_{5r_n}) \le \frac{\phi(5r_n) - \phi(r_n)}{\phi(5r_n) - \phi(0)} \le \frac{2^{-r_n}}{1 - 2^{-5}} \le 2^{1-r_n} \tag{8}$$

as $n \to \infty$ which proves the desired result. □

Remark 5. As noted after the construction, when a collision occurs it will connect a vertex at distance k with one at distance k or to one at distance $k+1$ that already has a neighbor at distance k. In the first case at distance k the comparison chain moves towards x with probability $\le 1/3$, the chain stays at the same distance with probability $\le 1/3$ and moves further away with probability $\ge 1/3$. In the second case at distance $k+1$ the comparison chain moves toward the root with probability $\le 2/3$ and further away with probability $\ge 1/3$.

If we have a birth and death chain X_n that jumps $p(k, k+1) = p_k$, $p(k,k) = r_k$ and $p(k, k-1) = q_k$ then

$$\phi(k+1) - \phi(k) = \frac{q_k}{p_k}[\phi(k) - \phi(k-1)]$$

recursively defines a function ϕ so that $\phi(X_n)$ is a martingale. In our comparison chain $q_k/p_k = 1/2$ for all but one value of k, which has $q_k/p_k \le 1$, so calculations like the one in (8) will work but give a slightly larger constant. Because of this, we will suppress these annoying details by assuming the graph is exactly a tree up to distance $(1/5)\log n$.

To prepare for the next result we need

Lemma 2.2. *If S_k is the sum of k independent mean one exponentials then*

$$P(S_k \le ak) \le \left(\frac{ae}{1+a}\right)^k$$

Remark 6. This holds for all a but is only useful when $ae/(1+a) < 1$, which holds if $a < 1/2$.

Proof. Let $\theta > 0$ and note $\int_0^\infty e^{-\theta x} e^{-x}\,dx = 1/(1+\theta)$. Using Markov's inequality we have

$$e^{-\theta ak} P(S_k \le ak) \le (1+\theta)^{-k}$$

Taking $\theta = 1/a$ and rearranging gives the desired result. □

Lemma 2.3. *Suppose two particles are a distance $r_n = 2\log_2 \log n$. Then with high probability the two particles will not collide by time $\log^2 n$.*

Proof. By Lemma 2.1 the probability of hitting $5r_n$ before 0 starting from r_n is

$$\ge 1 - 2/(\log n)^2 \tag{9}$$

A particle must make $4r_n$ jumps to go from distance $5r_n$ to r_n. Since jumps occur at rate 1 in the site model and at rate $\le M$ in the edge model, the last lemma implies that the probability of $k = r_n$ jumps in time $\le ar_n/M$ is

$$\le (ae)^{r_n} \le 1/(\log^3 n)$$

for large n if a is small enough. If a particle makes $2M(\log^2 n)/ar_n$ attempts to reach 0 before $5r_n$ starting from r_n then (9) implies that with high probability it will not be successful, while the last bound implies that this number of attempts will take time $\ge 2\log^2 n$ with high probability. □

Lemma 2.4. *Suppose two particles are a distance $r_n = 2\log_2 \log n$ and let $s_n/n \to 0$. Then with high probability the two particles will not hit by time s_n.*

Proof. Lemma 2.3 takes care of times up to $\log^2 n$. The result in (7) implies that if n is large then for $t \ge \log^2 n$, $p_t(x,y) \le 2/n$. Summing we see that if the two particle move independently the expected amount of time the two particles spend at the same site at times in $[\log^2 n, s_n]$ is $\le 2s_n/n \to 0$. Since the jump rates are bounded above this implies the desired result. □

In the proof of Lemma 3.10 we need the following result for hitting times.

Lemma 2.5. *Let $L = (1/5)\log_M n$. Suppose two particles performing independent continuous time (site or edge) random walks start at points are separated by distance k. Then there is a constant so that the probability the two particles hit by time $C_1 \log n$ is $\le C2^{-(k\wedge L/2)}$.*

Proof. Using the bound in Lemma 2.2 as in the proof of Lemma 2.3, if we replace C_1 by $C_2 = 4MC_1$ then we can treat the discrete time random walk in which on each step, one of the particles is picked at random to jump. L is chosen so that arguments in Section 2.2 show that when either particle looks at the ball of radius L around them, what they see differs from like a tree by at most one edge. If we ignore the extra edge, which is justified by

Remark 5, and if we let D_n be the distance between the two particles after n jumps, then (8) implies that if $k \le L/2$

$$P_k(T_0 < T_L) \le \frac{2^{-k} - 2^{-L}}{1 - 2^{-L}} \le 2 \cdot 2^{-k}$$

if n is large.

Suppose now that the distance between the two points is $\ge L/2$. The particles cannot hit until they first reach a distance of $L/2$, at which point the previous estimate can be applied. The journey from $k \le L/2$ to L takes at least $L/2$ steps. Thus if a particle makes K cycles from $L/2$ to L it has used up $KL/2$ units of time which is larger than $C_2 \log n$ if K is chosen large enough. This shows that the estimate holds with $C = 2K$. □

2.4 Results for the dual process

In this section we will consider the dual process on its original time scale, i.e., jumps occur at rate $O(1)$. In either version of the model, the rate at which branching occurs is $\le L/\lambda$ where $L = M^2$. (Here we are using the fact in the edge model the degree is bounded.) Let R_n be time of the nth branching. If $t_n = c_2 \log n$ for some constant $c_2 > 0$ then

$$P(R_1 \le t_n) \to 0 \qquad as\ n \to \infty$$

Let $N(t)$ be the number of branching events by time λt. Comparing with a branching process we have $EN(t) \le e^{Lt}$. The expected number of branchings in the interval $[\lambda t - t_n, \lambda t]$ is $\le e^{Lt}(c_2 \log n)/\lambda$ so as $n \to \infty$,

$$P(\lambda t - R_{N(t)} \le t_n) \to 0 \tag{10}$$

In the next three results C_1 is the constant defined in (7) and we make the following assumption:

($\star$) Suppose there are k particles in the dual at time 0, and each pair are separated by a distance $r_n = 2 \log_2 \log n$.

Lemma 2.6. *Suppose that at time 0, the first particle encounters an branching event. By time $C_1 \log n$, there may be coalescences between new born particles or with their parent, but with high probability there will be no other coalescences.*

Proof. This follows from Lemma 2.3. □

Lemma 2.7. *At time $C_1 \log n$ all the particles are almost uniformly distributed on the graph with the bound on the total variation distance uniform over all configurations allowed by ($\star$).*

Proof. This follows from (7). □

Lemma 2.8. *After time $C_1 \log n$, with high probability there is no coalescence between particles before the next branching event, and right before the next branching event, all the particles are r_n apart away from each other.*

Proof. The claim about coalescence follows from Lemma 2.4. The branching time is random but it is independent of the movement of the particles, so the result about the separation between particles follows from (7). □

Together with (10), Lemma 2.8 implies that there is no coalescence in the dual $[R_{N(t)}, \lambda t]$ and particles are at least r_n apart right before $R_{N(t)}$. According to Lemma 2.7, the coalescences between new born particles and their parents can only happen before $R_{N(t)} + C_1 \log n$, with no other coalescences. Lemma 2.7 tells us at times $\geq R_{N(t)} + C_1 \log n$, all the particles are almost uniformly distributed over the graph. Thus when we feed values into the dual process to begin to compute the state of x at time t the values are independent and equal to 1 with probability u.

Lemma 2.9. *$EU^n(t)$ converges to a limit $u(t)$.*

Proof. Let $Z(s)$, $s \leq t$ be the number of particles in the dual process, when we impose the rule that the number of particles is not increased until time $(C_1 \log n)/\lambda$ after a branching event. Our results imply that $Z(s)$ converges to a branching process. The last result shows that when we use the dual to compute the state of x at time t we put independent and identically distributed values at the $Z(t)$ sites. The result now follows from results in [5]. □

Lemma 2.10. *$U^n(t) - EU^n(t)$ converges in probability to 0.*

Proof. It follows from Lemma 2.4 that if $|x-y| > r_n$ then there will be no collisions between particles in the dual processes starting from x and y, and hence the values we compute for x and y are independent. The result now follows from Chebyshev's inequality. □

2.5 Computation of the reaction term

The final step is to show that $u(t)$ satisfies the differential equation. To warm up for the real proof, we begin by doing this on $\mathbb{Z}^d$ If ν_u is the voter model stationary distribution with density u and v_1 and v_2 are randomly chosen neighbors of x then

$$\langle h_{1,2}(x,\xi)\rangle_u = \nu_u(\xi(x) = 1, \xi(v_1) = 2 \text{ or } \xi(v_2) = 2)$$

The right-hand side can be computed using the duality between the voter model and coalescing random walk. Following the approach in Section 4 of [10] if we let $p(x|y|z)$ be the probability the random walks starting from x, y, and z never hit and $p(x|y,z)$ be the probability y and z coalesce but don't hit x then

$$\nu_u(\xi(x) = 1, \xi(y) = 2 \text{ or } \xi(z) = 2) = p(x|y|z)u(1-u^2) + q(x,y,z)u(1-u)$$

where $q(x,y,z) = p(x|y,z) + p(x,y|z) + p(x,z|y)$

Using this identity we can compute the reaction term defined in (1)

$$\begin{aligned}
\phi(u) &= \langle h_{2,1}(x,\xi) - h_{1,2}(x,\xi)\rangle_u \\
&= p(x|v_1|v_2)(1-u)(1-(1-u)^2) + q(x,v_1,v_2)u(1-u) \\
&\qquad - [p(x|v_1|v_2)u(1-u^2) + q(x,v_1,v_2)u(1-u)] \\
&= p(x|v_1|v_2)[(1-u)u(2-u) - u(1-u)(1+u)] \\
&= p(x|v_1|v_2)u(1-u)(1-2u)
\end{aligned} \tag{11}$$

The computations for the random graph are similar but in that setting we have to take into account the degree of x and what the graph looks like locally seen from x. Let q_k be the size-biased distribution kp_k/μ where $\mu = \sum_k p_k$ is the mean degree. Let $\mathbb{P}_k$ be a Galton Watson tree in which the root has degree k and the other vertices have j children with probability q_{j+1}.

In the site version a dual random walk path will spend a fraction $\pi^s(k) = q_k$ at vertices with degree k so

$$\langle h^s_{2,1} - h^s_{1,2} \rangle_u = \sum_k q_k \mathbb{P}_k(x|v_1|v_2) u(1-u)(1-2u)$$

where v_1 and v_2 are randomly chosen neighbors of the root. In the edge version $\pi^e(k) = p_k$ so

$$\langle h^e_{2,1} - h^e_{1,2} \rangle_u = \sum_k p_k \mathbb{P}_k(x|y|z) u(1-u)(1-2u)$$

3 Proof of Theorem 3

Recall that the density in the time-rescaled latent voter model is given by:

$$X_t = \sum_{x \in G_n} \pi(x) 1_{(\xi_{\lambda t}(x)=1)}. \tag{12}$$

To complete the proof of Theorem 3 using the result of Darling and Norris [7] given in Theorem 4 we need to estimate the probability of

$$\Omega_1^c = \left\{ \int_0^t |\beta(X_s) - b(X_s)|\, ds > \eta \right\} \tag{13}$$

where $\beta(\xi) = \sum_{\xi' \neq \xi} (x(\xi') - x(\xi)) q(\xi, \xi')$ is the drift in the particle system and $b(u) = c_p u(1-u)(1-2u)$ is the drift in the ODE.

To begin to do this, we define $\tilde{\xi}_s$ to be the same as ξ_s for time $s \le \lambda t - C_1 \log n$, while on the time interval $(\lambda t - C_1 \log n, \lambda t]$, $\tilde{\xi}$ only follows the paths from voter events of ξ, ignoring those from branching events. Let

$$\tilde{X}_t = \sum_{x \in G_n} \pi(x) 1_{\{\tilde{\xi}_{\lambda t}(x)=1\}}$$

be the density of this new process $\tilde{\xi}$. In order to determine $\tilde{\xi}_{\lambda t}$, we run the coalescing random walks backward in time, starting from time λt and stopping at time $\lambda t - C_1 \log n$.

We will first outline the proof then go back and fill in the details. In Section 3.1 we will prove

Lemma 3.1. *For any $\varepsilon > 0$ and $m < \infty$ the probability that more than εn sites are changed by branching arrow is $\le n^{-m}$ for large n.*

Let $\tilde{u} = x(\xi_{\lambda t - C_1 \log n})$. To bound $P(\Omega_1^c)$ we will prove:

Lemma 3.2. *Suppose* $\log n \ll \lambda_n \ll n$ *and* $m > 0$. *There is a* $\delta > 0$ *independent of* m *and constants* C_m *so that for any* $\varepsilon > 0$ *if* $n \geq n_0(m)$

$$P\left(|\tilde{X}_t - \tilde{u}| > \epsilon | \mathcal{F}_{t-(C_1 \log n)/\lambda}\right) \leq \frac{C_m}{\varepsilon^{2m} n^{\delta m}} \tag{14}$$

This will follow from Chebyshev's inequality once we have a suitable estimate on the $2m$th moment (see Lemma 3.3 below). Using (12) we have

$$X_t - \tilde{u} = \sum_{x \in G_n} \pi(x)[1_{(\xi_{\lambda t}(x)=1)} - \tilde{u}],$$

so if we let $Y(x) = 1_{(\xi_{\lambda t}(x)=1)} - \tilde{u}$ then

$$E(X_t - \tilde{u})^{2m} = \sum_{x_1, \ldots x_{2m}} \pi(x_1) \cdots \pi(x_{2m}) \cdot Y(x_1) \cdots Y(x_{2m}).$$

We will use π^k to denote the distribution $\pi \times \cdots \times \pi$ on G_n^k. If we introduce the dual coalescing random walks $W_1, \ldots, W_{2m}$ starting from distribution π^{2m} and let $r = C_1 \log n$ then we can write this as

$$E[Y(W_1(r)) \ldots Y(W_{2m}(r))]$$

To estimate probabilities for coalescing random walks we introduce independent random walks $W_1', \ldots, W_{2m}'$ starting from distribution π^{2m}, and use these to construct the $W_1, \ldots, W_{2m}$ by dropping the higher number after collisions. To simplify notation let $Z_i = Y(W_i(C_1 \log n))$ and $Z_i' = Y(W_i'(C_1 \log n))$.

Lemma 3.3. *There is* $\delta > 0$ *so that for each* m *we have*

$$|E[Z_1 \ldots Z_{2m}]| \leq C_m n^{-\delta m}$$

This will be proved in Section 3.3 after we bound the coalescence probabilities in Section 3.2.

3.1 Ignoring branching

To prove Lemma 3.1 we begin by noting that since the branching rate is $1/\lambda$ we can suppose without loss of generality that $\lambda \leq n^{1/2}$. In order for a site to be changed a branching arrow must hit the dual process for the site but one branching arrow can change multiple sites. Consider the coalescing random walk starting from all sites occupied and let T_k be the amount of space time in $[0, C_1 \log n]$ occupied by k-particles, i.e., a particle that is a coalescence of k particles. Clearly

$$\sum_{k=1}^{\infty} k T_k = C_1 n \log n \tag{15}$$

Let Π_k be the number of branching arrows that hit k-particles. A standard large deviations result (see e.g., (2.6.2) and Exercise 3.1.4 in [9]) shows that there is a constant c_2 so that if $Z_\mu = \text{Poisson}(\mu)$ then

$$P(Z_\mu > 2\mu) \leq \exp(-C_2 \mu)$$

If $T_k \geq n^{2/3}$ then $P(\Pi_k > 2T_k/\lambda) \leq \exp(-C_2 n^{2/3}/\lambda)$, so if $\mathcal{K} = \{k : T_k \geq n^{2/3}\}$ then using (15)

$$P\left(\sum_{k\in\mathcal{K}} k\Pi_k > 2C_1(n\log n)/\lambda\right) \leq |\mathcal{K}| \exp(-C_2 n^{1/6})$$

To control the contribution from particles of large weight, we will get a bound on the largest particle weight seen. Let $N_x(s)$ be the size of the cluster containing the particle that started at x at time t when we run the coalescing random walk to time $t-s$ and let $N_{max}(s)$ be the size of the largest cluster. We will show

Lemma 3.4. *If $\alpha > 0$ and $m < \infty$ and $t = C_1 \log n$ then for large n*

$$P(N_{max}(t) > n^\alpha) \leq n^{-m}.$$

When $k \notin \mathcal{K}$ monotonicity implies $P(\Pi_k > 2n^{2/3}/\lambda) \leq \exp(-c_2 n^{1/6})$, so if n is large

$$P\left(\sum_{k\notin\mathcal{K}} k\Pi_k > n^{2\alpha} n^{2/3}/\lambda\right) \leq n^{-m} + n^\alpha \exp(-c_2 n^{1/6})$$

which proves Lemma 3.1. It remains then to prove Lemma 3.4.

We begin by considering the edge model.

Lemma 3.5. *If $s \geq 1/2M$ then $\mathbb{E}(N_x(s) - 1) \leq 4Mes$.*

Proof. Let $y \neq x$ and W^y be the edge random walk starting from y. Noting that when W^x and W^y hit, they stay together for a time $\geq 1/2M$ with probability e^{-1} gives

$$\mathbb{P}(W^x \textit{ and } W^y \textit{ hit by time } s) \times \frac{1}{2Me} \leq \int_0^{s+1/2M} \sum_z p_r(x,z)p_r(y,z)\,dr$$

Since the edge random walks are reversible with respect to the uniform distribution, the transition probability is symmetric

$$\int_0^{s+1/2M} \sum_z p_r(x,z)p_r(y,z)\,dr = \int_0^{s+1/2M} \sum_z p_r(x,z)p_r(z,y)\,dr \tag{16}$$

$$= \int_0^{s+1/2M} p_{2r}(x,y)\,dr \tag{17}$$

Using this we have

$$EN_x(s) = \sum_y \mathbb{P}(W^x \textit{ and } W^y \textit{ hit by time } s) \leq 2Me \int_0^{s+1/2M} dr \leq 4Mes$$

where in the last step we have used $s \geq 1/2M$ ∎

Our next step is to bound the second moment of $N_x(t)$.

Lemma 3.6. *If $s \geq 1/2M$ then $\mathbb{E}(N_x(s)-1)(N_x(s)-2) \leq 3(4Mes)^2$.*

Proof. We begin by observing that

$$\mathbb{E}(N_x(s)-1)(N_x(s)-2) = \sum_{x_1,x_2} P(x_1, x_2 \in N_x(s)).$$

where the sum is over $x_i \neq x$ and $x_1 \neq x_2$. We first consider the case in which x and x_1 are the first to collide, and we bound

$$\sum_{x_1,x_2,y,z} \int_0^{s+1/2M} p_r(x,y)p_r(x_1,y)p_r(x_2,z)P(z \in N_{y,r}(s))\,dr$$

where $N_{y,r}(s)$ is the cluster at time s of the random walk that starts at y at time r. As in the previous proof $2Me$ times this quantity will bound the desired hitting probability. By symmetry $\sum_{x_2} p(x_2,z) = \sum_{x_2} p(z,x_2) = 1$. Using Lemma 3.5

$$\sum_z P(z \in N_{y,r}(s)) \leq 4Mes$$

Using reversibility we can write what remains of the sum as

$$\sum_{x_1,y} \int_0^{s+1/2M} p_r(x,y)p_r(y,x_1)\,dr = \sum_{x_1} \int_0^{s+1/2M} p_{2r}(x,x_1)\,dr \leq 2s \tag{18}$$

The second case to consider is when x_1 and x_2 are the first to collide, and we bound

$$\sum_{x_1,x_2,y,z} \int_0^{s+1/2M} p_r(x_1,y)p_r(x_2,y)p_r(x,z)P(z \in N_{y,r}(s))\,dr$$

Using symmetry $p_r(x_1,y)p_r(x_2,y) = p_r(y,x_1)p_r(y,x_2)$ then summing over x_1, x_2 we have

$$\leq \sum_{y,z} \int_0^{s+1/2M} p_r(x,z)P(z \in N_{y,r}(s))\,dr$$

We have $P(z \in N_{y,r}(s)) = P(y \in N_{z,r}(s))$ because either event says y and z coalesce in $[r,s]$, so summing over y and using Lemma 3.5 the above is

$$\leq (4Mes) \sum_z \int_0^{s+1/2M} p_r(x,z)\,dr \leq (4Mes) \cdot 2s \tag{19}$$

Combining our calculations proves the desired result. □

Lemma 3.7. *If $s \geq 1/2M$ then $\mathbb{E}[(N_x(s)-1)\cdots(N_k(s)-k)] \leq C_k(4Mes)^k$ and hence*

$$\mathbb{E}N_x^m(s) \leq C_{m,M}(1+s)^m$$

Proof. The second result follows easily from the first since

$$x^m = 1 + \sum_{k=1}^{m} c_{m,k}(x-1)\cdots(x-k)$$

The first case is

$$\sum_{\substack{x_1,\dots,x_k,\\ y,z_1,\dots z_{k-1}}} \int_0^{s+1/2M} p_r(x,y)p_r(x_1,y)p_r(x_2,z_1)\dots p_r(x_k,z_{k-1})P(z_1,\dots z_{k-1}\in N_{y,r}(s))\,dr$$

Using symmetry and summing over $x_2,\dots,x_k$ removes the $p_r(x_2,z_1)\dots p_r(x_k,z_{k-1})$ from the sum. Next we sum over $z_1,\dots z_{k-1}$ (which are distinct) and use induction to bound the sum by $C_{k-1}(4Mes)^{k-1}$. Finally we finish up by applying (18).

The second case is

$$\sum_{\substack{x_1,\dots,x_k,\\ y,z_1,\dots z_{k-1}}} \int_0^{s+1/2M} p_r(x_1,y)p_r(x_2,y)p_r(x_3,z_1)\dots p_r(x_k,z_{k-2})$$

$$p_r(x,z_{k-1})P(z_1,\dots z_{k-1}\in N_{y,r}(s))\,dr$$

Using symmetry and summing over $x_1,\dots,x_k$ removes the

$$p_r(x_1,y)p_r(x_2,y)p_r(x_3,z_1)\dots p_r(x_k,z_{k-2}).$$

As in the previous proof $P(z_1,\dots z_{k-1}\in N_{y,r}(s)) = P(z_1,\dots z_{k-2},y\in N_{z_{k-1},r}(s))$, so summing over $z_1,\dots,z_{k-2},y$ and using induction we can bound the sum by $C_{k-1}(4Mes)^{k-1}$. Finally we finish up by applying (19) with $z=z_{k-1}$ □

Remark 7. To extend to the site case where we do not have symmetry, we note that reversibility of this model with respect to $\pi(y)=d(y)/D$ implies

$$p_r(y,z)\le d(y)p_r(y,z) = d(z)p_r(z,y)\le Mp_r(z,y)$$

so the proof works as before but we accumulate a factor of M each time we use symmetry.

Now we are ready to give an upper bound on the size of the maximal cluster $N_{max}(t)$ at time λt. Here and for the rest of the proof of Lemma 3.2, we only use moment bounds so the proof is the same for the edge and site models

Proof of Lemma 3.4. By Chebyshev's inequality

$$n^{\alpha k}P(N_x(t)>n^\alpha)\le C_{k,M}(1+t)^k$$

If we pick $k>(m+1)/\alpha$ then

$$P\left(\max_x N_x(t)>n^\alpha\right)\le \frac{n}{n^{k\alpha}}C_{k,M}(2\log^2 n)^k = o(n^{-m})$$

which proves the desired result. □

3.2 Bounds on coalescence probabilities

Recall that $W_1, \ldots, W_{2m}$ are coalescing random walks starting from distribution π^{2m}

Lemma 3.8. *Let H_{12} be the event that W_1 and W_2 hit by time $C_1 \log n$.*

$$P(H_{12}) \le (C \log n)/n.$$

Proof. Let $\Delta = \{(v,v) : 1 \le v \le n\} \subset G_n \times G_n$ be the "diagonal." Let W_1' and W_2' be independent random walks. Since $(W_1'(t), W_2'(t)) =_d \pi^2$ for all $t \ge 0$, the expected occupation time of Δ is $(C_1 \log n)\pi^2(\Delta)$. In the edge case π is uniform so $\pi^2(\Delta) = 1/n$. In the site case if we let $d(x)$ be the degree of x and $D = \sum_x d(x)$ then $\pi(x) = d(x)/D$ so

$$\pi^2(\Delta) = \sum_x \frac{d(x)^2}{D^2} \le \frac{M^2}{n}$$

since $d(x) \le M$, $D \ge n$, and $|G_n| = n$.

The jump rate for $(W_1'(t), W_2'(t))$ is 2 in the site case and $\le 2M$ in the edge case, so when W_1' and W_2' hit the expected time they spend together is $\ge 1/2M$, and we have

$$P(H_{12}) \le (C_1 \log n) \frac{M^2}{n} \cdot 2M$$

which proves the desired result. □

Remark 8. In what follows we will prove the result only for the site case, since the time change argument in the last paragraph of the proof can be used to extend the argument to the edge case.

The computation of higher order coalescence probabilities is made complicated by the fact that if particles 1 and 2 are the first to coalesce at time $T_{1,2}$ then the joint distribution of (W_1, W_3, W_4) at time T_{12} is not π^3. To avoid some of these difficulties, we will estimate the probability that coalescences occur in a specific pattern, and ignore collisions not consistent with the pattern. For example let $H_{12,34}$ be the event that

- 1 and 2 coalesce at T_{12}. We ignore collisions involving particles 3 and 4 before that time.
- At time $T_{12,34} > T_{12}$ particles 3 and 4 coalesce. We ignore particle 1 on $[T_{12}, T_{12,34}]$.

We say particle 1 in the second bullet because, in our coupling, at time T_{12} we drop W_2' and use W_1' to move the coalesced particle.

The act of ignoring particles may look odd but the reasoning is legitimate. Doing this enlarges the event that the coalescences occurred in the indicated pattern leading to an over estimate of the probabilities of interest.

Lemma 3.9. $P(H_{12,34}) \le C(\log^2 n)/n^2$.

Proof. By Lemma 3.8, the probability of the event H_{12} that 1 and 2 hit by time $C_1 \log n$ is $\le C(\log n)/n$. Since we are ignoring particles 3 and 4 up to time T_{12}, their joint distribution at time T_{12} conditional on $H_{1,2}$ is π^2. This implies that

$$P(H_{12,34}) \le P(H_{12}) \cdot (C \log n)/n$$

which proves the desired result. □

Consider now the event $H_{12,3}$

- 1 and 2 coalesce at time T_{12}. We ignore collisions involving particles 3 and 4 before that time.
- 3 coalesces with particle 1 at $T_{12,3} > T_{12}$. We ignore particle 4 during $[T_{12}, T_{12,3}]$.

Lemma 3.10. *There is an $\delta > 0$ so that $P(H_{12,3}) \le C \log n/n^{1+\delta}$*

Proof. As in the previous argument the probability that 1 and 2 hit is $\le C(\log n)/n$ and at time T_{12} the location of W_3' has distribution π and is independent of W_1'. Unfortunately W_1' does not have distribution π since it may be easier for particles 1 and 2 to coalesce at some points than others.

Let π_{min} be the minimal value of $\pi(x)$ and let $L = (1/5)\log_M n$. Using the observations that (i) since the max degree is $M \ge 3$, the number of vertices at distance k from a fixed vertex is $\le M^k$ and (ii) by Lemma 2.5 the probability two particles separated by $k \le L$ hit by time $C_1 \log n$ is $\le C2^{-k}$, and (iii) two particles that are separated by more that L must come to within distance L before they hit.

$$\begin{aligned} P(H_{12,3}|H_{12}) &\le \sum_{k=1}^{L/2} 2^{-k} M^k \pi_{min} + 2^{-L/2}(n - M^{L/2})\pi_{min} \\ &\le 2^{-L/2} n \pi_{min} \sum_{j=0}^{\infty} (2/M)^j \le \frac{C 2^{-(1/10)\log_M n}}{1 - 2/3} \le C n^{-\delta} \end{aligned}$$

which proves the desired result. □

Lemmas 3.9 and 3.10 contain all the ideas needed to estimate general coalescence patterns. The hardest part of doing this in general is to find appropriate notation to enumerate the possibilities. To do this we will use notation used to describe phylogenetic trees. For example

$$H_{(((12)5)9),(34),((67)8)}$$

means that first 1 and 2 coalesce, then 3 and 4 coalesce, next 5 coalesces with (12), 6 and 7 coalesce, 8 coalesces with them and finally 9 coalesces with ((12)5). In defining these events we ignore collisions between particles that have already coalesced with another one. In the example under consideration we have

$$P(H_{(((12)5)9),(34),((67)8)}) \le C \left(\frac{\log n}{n}\right)^3 \cdot n^{-3\delta} \le C n^{-9\delta/2}$$

We say a random walk W_i is *isolated* if W_i does not coalesce with other random walks by time $C_1 \log n$. Suppose s is the number of isolated random walks among $W_1, \dots W_{2m}$ and define

$$G^{NI}_{i_1,\dots,i_k} = \{\text{None of } W_{i_1}, \dots, W_{i_k} \text{is isolated}\}$$
$$G^{iso}_{i_1,\dots,i_k} = \{\text{All } W_{i_1}, \dots, W_{i_k} \text{ are isolated}\}$$

Since $G^{NI}_{i_1,\dots,i_k}$ is contained in a union of H events involving k particles.

Lemma 3.11. *Given any k coalescing random walks $W_{i_1}, \dots, W_{i_k}$ starting from the stationary distribution, then there exist constants $C > 0$ and $\delta > 0$ such that the probability of no isolated random walk is bounded by*

$$P\left(G^{NI}_{i_1 \dots i_k}\right) \le C/n^{\delta k/2} \tag{20}$$

3.3 Moment estimates

We begin with second moments.

Lemma 3.12. *If n is large $E[Z_1 Z_2] \le (C \log n)/n$.*

Proof. On H_{12},

$$Y(W_1(r))Y(W_2(r)) = Y^2(Z_1(r)) \le 1$$

so the contribution to the expected value is $\le C(\log n)/n$. Since $W_1'(r)$ and $W_2'(r)$ are independent and have distribution π

$$E[Y(W_1'(r))Y(W_2'(r))] = 0$$

Using our coupling and this we get

$$\begin{aligned} |E[Y(W_1(r))Y(W_2(r)); H^c_{12}]| &= |E[Y(W_1'(r))Y(W_2'(r)); H^c_{12}]| \\ &= |E[Y(W_1'(r))Y(W_2'(r)); H_{12}]| \le P(H_{12}) \le \frac{C \log n}{n} \end{aligned}$$

and the desired result follows. □

Turning now to 4th moments, let S be the number of isolated random walks among W_1, W_2, W_3 and W_4. Then

$$E[Z_1 Z_2 Z_3 Z_4] = \sum_{s=1}^{4} E\left[Z_1 Z_2 Z_3 Z_4; S = s\right] \tag{21}$$

Case 1: s=1. First by symmetry, we have

$$E\left[Z_1 Z_2 Z_3 Z_4; S = 1\right] = 4E\left[Z_1 Z_2 Z_3 Z_4; G^{NI}_{123} \cap G^{iso}_4\right]$$

where 4 comes from the choices of the only isolated random walk. Now we couple W_4 to an independent random walk W_4'. Precisely, $(W_1, W_2, W_3, W_4) = (W_1, W_2, W_3, W_4')$ until

$W_4 = W_4'$ hits the trajectory of W_i for some $i \le 3$. When this occurs, the first vector becomes (W_1, W_2, W_3, W_i). From this coupling, we know

$$E\left[Z_1Z_2Z_3Z_4; G_{123}^{NI} \cap G_4^{iso}\right] = E\left[Z_1Z_2Z_3Z_4'; G_{123}^{NI} \cap G_4^{iso}\right]$$

Note that Z_4' is independent of Z_1, Z_2, Z_3. Hence $E\left[Z_1Z_2Z_3Z_4'; G_{123}^{NI}\right] = 0$. This implies

$$\begin{aligned} &\left|E\left[Z_1Z_2Z_3Z_4'; G_{123}^{NI} \cap G_4^{iso}\right]\right| = \left|E\left[Z_1Z_2Z_3Z_4'; G_{123}^{NI} \cap G_4^{iso,c}\right]\right| \\ \le& 3\left|E\left[Z_1Z_2Z_3Z_4'; G_{123}^{NI} \cap G_{14}\right]\right| \le 3P(G_{123}^{NI} \cap G_{14}) \le \frac{C}{n^\delta} P(H_{12,3}) \le C\log n/n^{1+2\delta} \end{aligned} \tag{22}$$

Case 2: s=2. Similarly, symmetry tells us

$$E\left[Z_1Z_2Z_3Z_4; S=2\right] = 6E\left[Z_1Z_2Z_3Z_4; G_{12}^{NI} \cap G_{34}^{iso}\right] = 6E\left[Z_1Z_2Z_3Z_4; G_{12} \cap G_{34}^{iso}\right]$$

Now couple (W_3, W_4) to independent random walks (W_3', W_4'). Precisely, (W_1, W_2, W_3, W_4) and (W_1, W_2, W_3', W_4') are the same until W_3 or W_4 hits others. Then

$$E\left[Z_1Z_2Z_3Z_4; G_{12} \cap G_{34}^{iso}\right] = E\left[Z_1Z_2Z_3'Z_4'; G_{12} \cap G_{34}^{iso}\right]$$

Note that Z_3' and Z_4' are independent of W_1 and W_2. Hence $E\left[Z_1Z_2Z_3'Z_4'; G_{12}\right] = 0$. This implies

$$\begin{aligned} &\left|E[Z_1Z_2Z_3'Z_4'; G_{12} \cap G_{34}^{iso}]\right| = \left|E[Z_1Z_2Z_3'Z_4'; G_{12} \cap G_{34}^{iso,c}]\right| \\ &\le C\left(P(H_{12,3}) + P(H_{12,34})\right) \le C\log n/n^{1+\delta} \end{aligned}$$

The first inequality holds because on G_{12} if W_3' is not isolated, it can either hit W_1 or W_2, which has probability bounded by $CP(H_{12,3})$ where the constant C takes care of the order of coalescent; or it can hit W_4' without hitting W_1 or W_2, which has probability bounded by $P(H_{12,34})$. The same argument applies to W_4'. Therefore, we have obtained

$$E\left[Z_1Z_2Z_3Z_4; S=2\right] \le C\log n/n^{1+\delta} \tag{23}$$

Case 3: s=3. Impossible

Case 4: s=4. Couple W_i, $1 \le i \le 4$ to independent random walks W_i', $1 \le i \le 4$. Precisely, they agree until there is a coalescence. Then

$$\begin{aligned} &E\left[Z_1Z_2Z_3Z_4; S=4\right] = E\left[Z_1Z_2Z_3Z_4; G_{1234}^{iso}\right] \\ &= E\left[Z_1'Z_2'Z_3'Z_4'; G_{1234}^{iso}\right] = -E\left[Z_1'Z_2'Z_3'Z_4'; G_{1234}^{iso,c}\right] \end{aligned}$$

Now $G_{1234}^{iso,c}$ is a disjoint union of events

$$G_{A^c}^{NI} \cap G_A^{iso}$$

where $A \subsetneqq \{1,2,3,4\}$ and $A^c = \{1,2,3,4\}\backslash A$. Combining this with (21), (22), and (23) gives Lemma 3.3 for $m = 2$.

General m. To compute $E[Z_1 Z_2 \dots Z_{2m}]$, let S denote the number of isolated random walks among $W_1, \dots, W_{2m}$ at time $C_1 \log n$. Then

$$E[Z_1 Z_2 \dots Z_{2m}] = \sum_{s=0}^{2m} E[Z_1 Z_2 \dots Z_{2m}; S = s] \tag{24}$$

For any $1 \le s \le 2m$, first symmetry gives us

$$E[Z_1 Z_2 \dots Z_{2m}; S = s] = \binom{2m}{s} E[Z_1 Z_2 \dots Z_{2m}; G^{NI}_{1\dots 2m-s} \cap G^{iso}_{2m-s+1\dots 2m}] \tag{25}$$

Hence we just need to focus on the case where the last s random walks are isolated while no isolated random walk appears in the first $2m-s$ random walks. Now couple W_j with $2m-s+1 \le j \le 2m$ to independent random walks W'_j with $2m-s+1 \le j \le 2m$. Precisely, $(W_1, \dots, W_{2m})$ and $(W_1, \dots, W_{2m-s}, W'_{2m-s+1}, \dots, W'_{2m})$ are identical until any W_i with $2m-s < i \le 2m$ hits another particle. Then

$$\begin{aligned} I :=& E\left[Z_1 \dots Z_{2m}; G^{NI}_{1\dots 2m-s} \cap G^{iso}_{2m-s+1\dots 2m}\right] \\ =& E\left[Z_1 \dots Z_{2m-s} Z'_{2m-s+1} \dots Z'_{2m}; G^{NI}_{1\dots 2m-s} \cap G^{iso}_{2m-s+1\dots 2m}\right] \end{aligned} \tag{26}$$

As before, note that $Z'_{2m-s+1}, \dots, Z'_{2m}$ are all independent of $Z_1, \dots, Z_{2m-s}$. This implies that $E\left[Z_1 \dots Z_{2m-s} Z'_{2m-s+1} \dots Z'_{2m}; G^{NI}_{1\dots 2m-s}\right] = 0$ and that

$$E[Z_1 \dots Z'_{2m}; G^{NI}_{1\dots 2m-s} \cap G^{iso}_{2m-s+1\dots 2m}] = -E[Z_1 \dots Z'_{2m}; G^{NI}_{1\dots 2m-s} \cap G^{iso,c}_{2m-s+1\dots 2m}] \tag{27}$$

Note that $G^{NI}_{1\dots 2m-s} \cap G^{iso,c}_{2m-s+1\dots 2m}$ is a union of disjoint sets of the form

$$G^{NI}_{A^c} \cap G^{iso}_{A}$$

where $A \subsetneq \{2m-s+1, \dots, 2m\}$ and $A^c = \{1, 2, \dots, 2m\} \backslash A$. Combining this with (26) and (27) gives

$$I \le \sum_{A \subsetneq \{2m-s+1, \dots, 2m\}} \left| E[Z_1 \dots Z'_{2m}; G^{NI}_{A^c} \cap G^{iso}_{A}] \right|$$

Since $|A| \le s-1$, we have reduced the number of isolated random walks by at least 1. Now apply similar argument to each $G^{NI}_{A^c} \cap G^{iso}_{A}$. Note that $Z'_i,\ i \in A$ are independent of the Z_i $i \in A^c$. Hence $E[Z_1 \dots Z'_{2m}; G^{NI}_{A^c}] = 0$. This implies that

$$E[Z_1 \dots Z'_{2m}; G^{NI}_{A^c} \cap G^{iso}_{A}] = -E[Z_1 \dots Z'_{2m}; G^{NI}_{A^c} \cap G^{iso,c}_{A}]$$

We can further subdivide each $G^{NI}_{A^c} \cap G^{iso,c}_{A}$ into disjoint sets of the form $G^{NI}_{B^c} \cap G^{iso,c}_{B}$ where $B \subsetneq A \subsetneq \{1, \dots, 2m\}$ and thus $|B| \le s-2$. This leads to

$$I \le C \sum_{B \subsetneq \{2m-s+1, \dots, 2m\},\ |B| \le s-2} \left| E[Z_1 \dots Z'_{2m}; G^{NI}_{B^c} \cap G^{iso}_{B}] \right|$$

where the constant C takes care of the possible repetition of $G^{NI}_{B^c} \cap G^{iso}_{B}$ from those subdivisions. We can keep doing this and decrease the number of isolated random walks by at least

1 at each step until we have no isolated random walk, i.e. when all those A (or B)given above have $|A| = 0$ (or $|B| = 0$). We will eventually have

$$I \leq C \left| E[Z_1 \dots Z'_{2m}; G^{NI}_{1,2,\dots,2m}] \right|$$

Since $\left| Z_1 \dots Z_{2m-s} Z'_{2m-s+1} \dots Z'_{2m} \right| \leq 1$, then by Lemma 3.11 we have

$$I \leq CP \left(G^{NI}_{12\dots 2m} \right) \leq C/n^{\delta m} \tag{28}$$

and the proof of Lemma 3.3 is complete.

3.4 Bounding the drift

The drift

$$\beta(\xi_{\lambda t}) = \sum_{x \in G_n} \pi(x) \sum_{y \sim x} \sum_{z \sim x, z \neq y} [1_{\{\xi_{\lambda t}(x)=2,\ \xi_{\lambda t}(y)=1 \text{ or } \xi_{\lambda t}(z)=1\}} - 1_{\{\xi_{\lambda t}(x)=1,\ \xi_{\lambda t}(y)=2 \text{ or } \xi_{\lambda t}=2\}}]$$

We want to show

Lemma 3.13. *There is a $\gamma > 0$ so that for any m there is a constant C_m so that*

$$P(|\beta(\xi_{\lambda t}) - b(X_t)| \geq \epsilon | \mathcal{F}_{t-(C_1 \log n)/\lambda}) \leq \frac{C_m}{\varepsilon^{2m} n^{2m\gamma}}. \tag{29}$$

Proof. We prove the result for the edge case and leave the straightforward extension to the site case to the reader. If we let $\mathbb{1}(x|y|z)$ is the indicator function of the event that the dual random walks starting from x, y, and z at time t do not hit by time $t - C_1(\log n)/\lambda$ and $p(x|y|z) = E\mathbb{1}(x|y|z)$ then

$$E[\beta(\xi_t)|\mathcal{F}_{t-C_1(\log n)/\lambda}] = \frac{1}{n} \sum_{x \in G_n} \sum_{y \sim x} \sum_{z \sim x, z \neq y} \mathbb{1}(x|y|z)\tilde{u}(1-\tilde{u})(1-2\tilde{u}) \tag{30}$$

$$b(X_t) = \frac{1}{n} \sum_{x \in G_n} \sum_{y \sim x} \sum_{z \sim x, z \neq y} p(x|y|z)\tilde{u}(1-\tilde{u})(1-2\tilde{u}) \tag{31}$$

The random variables $\mathbb{1}(x|y|z)$ are dependent if the triples (x, y, x) and (x', y', z') overlap or if the associated random walks coalesce. To simplify things we will let $\hat{\mathbb{1}}(x|y|z)$ be the event none of the walks r-coalesce, i.e., the pair collides before either of them exits $B(x, r)$, where $r = \varepsilon \log n$ and ε will be chosen later to be small enough. Let

$$Y_{x,y,z} = \hat{\mathbb{1}}(x|y|z) - p(x|y|z) \quad \text{and} \quad \hat{Y}_{x,y,z} = \hat{\mathbb{1}}(x|y|z) - \hat{p}(x|y|z)$$

where $\hat{p}(x|y|z) = E(\hat{\mathbb{1}}(x|y|z))$. As before we will compute $2m$th moments of the sum over $x \in G_n$ and neighbors y, $z \neq y$ of x. The calculation is simpler here than in Sections 3.2–3.3, since we are only concerned whether the particles coalesce and not how they are spread over the graph at time $C_1 \log n$.

Lemma 3.14. *There is a $\beta > 0$ so that for any m we have*

$$E\left(\sum_{x,y,z} \hat{Y}_{x,y,z}\right)^{2m} \le C_m n^{1+\beta}.$$

Proof. The sum has $K = \sum_x d(x)(d(x)-1)$ terms. The $2m$th moment of the sum has terms of the form.

$$\hat{Y}_{x_1,y_1,z_1} \cdots \hat{Y}_{x_{2m},y_{2m},z_{2m}}$$

If some x_i has distance $3r$ from all of the other x_j then Y_{x_i,y_i,z_i} is independent of the product of the rest of the random variables and the expected value is 0.

Suppose now that for each x_i there is at least one x_j that is within distance $3r$. Create a graph D (for dependency) where there is an edge between i and j if $d(x_i, x_j) < 3r$. Let κ be the number of components in the graph. The number of points within distance $3r = 3\varepsilon \log n$ of a given x is $\le M^{3\varepsilon \log n} \equiv n^\beta$. If the dependency graph has κ components the number of terms $\le A_D n^\kappa n^{\beta(2m-\kappa)}$. When D has no singletons $\kappa \le m$. Since $E|Y_{x_1,y_1,z_1} \cdots Y_{x_{2m},y_{2m},z_{2m}}| \le 1$ the desired result follows. □

To bound the sum of the $Y_{x,y,z}$ we will write

$$Y_{x,y,z} = \hat{Y}_{x,y,z} + (\hat{p}(x|y|z) - p(x|y|z)) + (\mathbb{1}(x|y|z) - \hat{\mathbb{1}}(x|y|z))$$

To control the middle term note that Lemma 2.3 implies that with high probability the two walks that are separated by $r = \varepsilon \log n$ will not hit before they are separated by $5r$ is $\le 2^{1-\varepsilon \log n} \equiv 2n^{-\alpha}$. Using this result repeatedly we see the probability they do not hit by $C_1 \log n$ is $\le Cn^{-\alpha}$. Thus

$$\sum_{x,y,z} |p(x|y|z) - \hat{p}(x|y|z)| \le cn^{1-\alpha}$$

To control the $2m$th moment of the sum of the third term, suppose we are given $1 \le K \le 2m$ distinct (x_i, y_i, z_i) where y_i and z_i are different neighbors of x_i. Note that $Z = \prod_{i=1}^K \left(\mathbb{1}(x_i|y_i|z_i) - \hat{\mathbb{1}}(x_i|y_i|z_i)\right) > 0$ if and only if $\mathbb{1}(x_i|y_i|z_i) - \hat{\mathbb{1}}(x_i|y_i|z_i) > 0$ for all $1 \le i \le K$. That is, for any i, there exist a pair among $(W^{x_i}, W^{y_i}, W^{z_i})$ such that they do not $r-$coalesce but coalesce after exiting $B(x, r)$. As before, we only focus on coalescent in a specific pattern and consider the event H as the following:

- Suppose $T_0 = 0$. Some particles from $(W^{x_1}, W^{y_1}, W^{z_1})$ coalesce at time T_1 but do not $r-$coalesce. We ignore collisions involving other particles and let them do independent random walks before that time.

- $(W^{x_1}, W^{y_1}, W^{z_1})$ are ignored immediately after time T_1 except for the ones who appear in $(W^{x_2}, W^{y_2}, W^{z_2})$. At time $T_2 > T_1$, some particles from $(W^{x_2}, W^{y_2}, W^{z_2})$ coalesce after exiting $B(x_2, r)$. We ignore all other particles on $[T_1, T_2]$.

- In general, at time $T_k > T_{k-1}$, some particles from $(W^{x_k}, W^{y_k}, W^{z_k})$ coalesce after exiting $B(x_k, r)$. We ignore all other particles on $[T_{k-1}, T_k]$.

This again enlarges the probability of our interest up to a constant factor from permutation. Moreover by Lemma 2.5, each step has probability $\leq C/n^{-\alpha}$ to occur . Hence

$$\begin{aligned} P(Z>0) &\leq CP(H) \\ &\leq C\prod_{k=1}^{2m} P\left(T_k<\infty|T_{k-1}\infty\right) \leq C/n^{-K\alpha} \end{aligned}$$

Note that in the expansion of $\left(\sum_{x,y,z}\mathbb{1}(x|y|z)-\hat{\mathbb{1}}(x|y|z)\right)^{2m}$, the number of terms consisting of such K distinct (x_i,y_i,z_i) is $\leq n^K M(M-1)$.

This implies that

$$E\left(\sum_{x,y,z}\mathbb{1}(x|y|z)-\hat{\mathbb{1}}(x|y|z)\right)^{2m} \leq \sum_{K=1}^{2m} n^K M(M-1)\times C/n^{-K\alpha} \leq C_m n^{2m(1-\alpha)}.$$

Combining our results and using Chebyshev's inequality, the proof of Lemma 3.13 is complete.

3.5 Final details

To extend Lemma 3.13 to bound the probability of

$$\Omega_1^c=\left\{\int_0^t |\beta(X_s)-b(X_s)|\,ds \geq \eta\right\}$$

we subdivide the interval $[0,t]$ into subintervals of length $1/\lambda n^{1/2}$. Within each interval the probability that more than $2n^{1/2}$ sites will flip is $\leq \exp(-c\sqrt{n})$. From this it follows that if $2t\varepsilon\leq\eta$ then

$$P(\Omega_1^c)\leq t\lambda n^{1/2}\left[\frac{C_{m,\varepsilon}}{n^{m/2}}+\exp(-c\sqrt{n})\right] \tag{32}$$

The last bound only works for fixed t. To get long time survival we will iterate. Let

$$T_0=\inf\{t:|x_t-1/2|<\varepsilon\}$$

and note that x_t is the solution of the ODE so this is not random. this is not random. Theorem 4 implies that at this time $|X_t-1/2|\leq 2\varepsilon$ with very high probability, i.e., with an error of less that $Cn^{-(m-1)/2}$. Let

$$T_1=\inf\{t>T_0:|X_t-1/2|\geq 4\varepsilon\}$$

and note that on $[T_0,T_1]$ we have $|X_t-1/2|\leq 4\varepsilon$. There is a constant t_0 so that if $x(0)=1/2+4\varepsilon$ or $x(0)=1/2-4\varepsilon$ then $|x(t_0)-1/2|\leq\varepsilon$. Let $S_1=T_1+t_0$. Theorem 4 implies that with high probability $|X(S_1)-1/2|\leq 2\varepsilon$ and $|X_t-1/2|\leq 5\varepsilon$ on $[T_1,S_1]$. For $k\geq 2$ let

$$T_k=\inf\{t>S_{k-1}:|X_t-1/2|\geq 4\varepsilon\} \quad \text{and} \quad S_k=T_k+t_0.$$

We can with high probability iterate the construction $n^{(m-2)/2}$ times before it fails. Since each cycle takes at least t_0 units of time, the proof of Theorem 3 is complete.

Acknowledgments

Several years ago Shirshendu Chatterjee worked with the second author on this model on random regular graphs. Some of the ideas from that work are used here but many new ideas are need to deal with the fact that $d(x)$ is not constant. The authors would like to thank Eric Foxall for telling them about Darling and Norris' work, and thank two referees for comments that improved the readability of the paper. Both authors were partially supported by NSF grant DMS 1505215 from the probability program.

References

[1] Bramson, M., and Griffeath, D. (1980) Asymptotics for interacting particle systems on $\mathbb{Z}^d$. *Z. Wahrscheinlichkeitstheorie verw. Gebiete.* 53, 183–196

[2] Cooper, C., Frieze, A., and Radzik, T. (2010) Multiple random walks in random regualr graphs. *SIAM J. Discrete Math.* 23, 1738–1761

[3] Cooper, C., Elsässer, R., Ono, H., and Radzik, T. (2012) Colescing random walks and voting on graphs. arXiv:1204.4106

[4] Cox, J.T., and Durrett, R. (2016) Evolutionary games on the torus with weak selection. *Stoch. Proc. Appl.* 126, 2388-2409

[5] Cox, J.T., Durrett, R., and Perkins, E.A. (2013) Voter model perturbations and reaction diffusion equations. *Astérisque.* Volume 349. arXiv:1103.1676

[6] Cox J.T., and Greven A (1990) On the long term behavior of some finite particle systems. *Probab. Theory Rel. Fields* 85, 195–237 .

[7] Darling, R.W.R., and Norris, J.R. (2008) Differential equation approximation for Markov chains. *Probability Surveys.* 5, 37–79

[8] Durrett, R. (2006) *Random Graph Dynamics.* Cambridge U. Press

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

[10] Durrett, R. (2014) Spatial evolutionary games with small selection coefficients. *Electronic J. Probability.* 19, paper 121

[11] Durrett, R.; Neuhauser, C. (1994) Particle systems and reaction-diffusion equations. *Ann. Probab.* 22, 289333.

[12] Gkantsis, C., Mihail, M., and Saberi, A. (2003) Conductance and congestion in power law graphs. *Proceedings of the 2003 ACM SIGMETRICS international conference on measurement and modeling of computer systems*, 148–159

[13] Lambiotte R., Saramaki, J., and Blondel, V.D. (2009) Dynamics of Latent Voters. *Physical Review E.* 79, paper 046107

[14] Lieberman, E., Hauert, C., and Nowak, M.A. (2005) Evolutionary dynamics on graphs. *Nature.* 433, 312–316

[15] Liggett, T.M. (1985) *Interacting Particle Systems.* Springer-Verlag, New York

[16] Liggett, T.M. (1999) *Stochastic Interacting Systems: Contact, Voter and Exclusion Processes.* Springer, New York.

[17] Maruyama, T. (1970) The effective number of alleles in a subdivided population. *Theor. Pop. Biol.* 1, 273–306

[18] Mountford, T. S. (1993) A metastable result for the finite multidimensional contact process. *Canad. Math. Bull.* 36, no. 2, 216226.

[19] Mountford, T., Mourrat, J-C., Daniel Valesin, D., and Yao, Q. Exponential extinction time of the contact process on finite graphs. arXiv:1203.2972

[20] Oliveira, R.I. (2012) On the coalescence time for reversible random walks. *Transactions of the AMS.* 364, 2109–2128

[21] Oliveira, R.I. (2013) Mean-field conditions for coalescing random walks. *Ann. Probab.* 41, 3420–3461

[22] Sawyer, S. (1979) A limit theorem for patch sizes in a selectively-neutral migration model. *J. Appl. Prob.* 16, 482–495

[23] Suchecki, K., Eguuíluz, V.M. and Miguel, M.S. (2005) Conservation laws for the voter model in complex networks. *Europhysics Letters.* 69, 228–234