

# CUTOFF FOR THE NOISY VOTER MODEL

J. THEODORE COX, YUVAL PERES, AND JEFFREY E. STEIF

ABSTRACT. Given a continuous time Markov Chain  $\{q(x, y)\}$  on a finite set  $S$ , the associated noisy voter model is the continuous time Markov chain on  $\{0, 1\}^S$  which evolves by (1) for each two sites  $x$  and  $y$  in  $S$ , the state at site  $x$  changes to the value of the state at site  $y$  at rate  $q(x, y)$  and (2) each site rerandomizes its state at rate 1. We show that if there is a uniform bound on the rates  $\{q(x, y)\}$  and the corresponding stationary distributions are “almost” uniform, then the mixing time has a sharp cutoff at time  $\log |S|/2$  with a window of order 1. Lubetzky and Sly proved cutoff with a window of order 1 for the stochastic Ising model on toroids: we obtain the special case of their result for the cycle as a consequence of our result. Finally, we consider the model on a star and demonstrate the surprising phenomenon that the time it takes for the chain started at all ones to become close in total variation to the chain started at all zeros is of smaller order than the mixing time.

## 1. INTRODUCTION

Consider a continuous time Markov chain on the finite set  $S$ ,  $|S| \geq 2$ , where the rate of going from  $x$  to  $y$  is  $q(x, y)$ . We let  $q_{\max} := \max\{\sum_{y \neq x} q(x, y) : x \in S\}$  be the maximum rate that we leave a state.

Next,  $(S, q)$  yields a continuous time Markov process on  $\{0, 1\}^S$  called *the noisy voter model with voting mechanism  $(S, q)$*  (often abbreviated *the noisy voter model*) where, independently, (1) for each two sites  $x$  and  $y$ , the state at site  $x$  changes to the value of the state at site  $y$  at rate  $q(x, y)$  and (2) each site rerandomizes its state at rate 1. Rerandomizes means that the state at that site switches to 1 or 0 each with probability 1/2 independently of everything else. The noisy voter model was introduced in [5] by Granovsky and Madras. Denoting an element of  $\{0, 1\}^S$  by  $\eta = \{\eta(x)\}_{x \in S}$ , one can describe this dynamic as saying: Independently at each  $x \in S$ ,

$$(1.1) \quad \begin{aligned} 0 &\rightarrow 1 \text{ at rate } \frac{1}{2} + \sum_{y \neq x} q(x, y)\eta(y) && \text{if } \eta(x) = 0, \\ 1 &\rightarrow 0 \text{ at rate } \frac{1}{2} + \sum_{y \neq x} q(x, y)(1 - \eta(y)) && \text{if } \eta(x) = 1. \end{aligned}$$

---

1991 *Mathematics Subject Classification.* 60J27,60K35.

*Key words and phrases.* noisy voter models, mixing times for Markov chains, cutoff phenomena.

Observe that whether or not  $(S, q)$  is irreducible, the corresponding noisy voter model is clearly irreducible, and hence has a unique stationary distribution. If there were no “rerandomization”, this would simply be the ordinary voter model associated to  $q$  which has, in the case that  $q$  is irreducible, two absorbing states, all 0’s and all 1’s. On the other hand, if there were no voter mechanism (essentially meaning that  $q(x, y) = 0$  for all  $x$  and  $y$ ), then the model would simply be continuous time random walk on the hypercube.

Throughout this paper, given  $q$ , we let  $\{\eta_t\}_{t \geq 0}$  denote the corresponding noisy voter model,  $\mu_\infty$  denote its stationary distribution and  $\mu_t^\eta$  denote the law of  $\eta_t$  when  $\eta_0 \equiv \eta$ . (The dependence of these on  $q$  is implicit.) If we have a sequence of such systems, we let  $\{\eta_t^n\}_{t \geq 0}$ ,  $\mu_\infty^n$  and  $\mu_t^{n, \eta}$  denote these objects for the  $n$ th system.

Recall that the total variation distance between two probability measures  $m_1$  and  $m_2$  on a finite set  $\Omega$  is defined to be

$$\|m_1 - m_2\|_{\text{TV}} := \frac{1}{2} \sum_{s \in \Omega} |m_1(s) - m_2(s)|.$$

Next, given a noisy voter model, for  $\varepsilon > 0$ , we let

$$t_{\text{mix}}(\varepsilon) := \inf\{t \geq 0 : \max_{\eta \in \{0,1\}^S} \|\mu_t^\eta - \mu_\infty\|_{\text{TV}} \leq \varepsilon\}$$

denote the  $\varepsilon$ -mixing time.

The main theorem of the paper is the following.

**Theorem 1.** *Assume that we have a sequence  $(S^n, q^n)$  of continuous time Markov chains with  $\lim_{n \rightarrow \infty} |S^n| = \infty$  and  $\sup_n q_{\text{max}}^n < \infty$ . Assume further that there is  $C$  such that for each  $n$ , there is a stationary distribution for  $(S^n, q^n)$  where the ratio of the largest and smallest point masses is at most  $C$  (this holds for example in any transitive situation). Then, for each  $\varepsilon$ ,*

$$(1.2) \quad t_{\text{mix}}(\varepsilon) = \frac{1}{2} \log |S^n| (1 + o(1)).$$

Moreover, we have that

$$(1.3) \quad \lim_{\alpha \rightarrow \infty} \liminf_{n \rightarrow \infty} \|\mu_{\frac{1}{2} \log |S^n| - \alpha}^{n, \mathbf{1}} - \mu_\infty^n\|_{\text{TV}} = 1$$

where  $\mathbf{1}$  denotes the configuration of all 1’s and

$$(1.4) \quad \lim_{\alpha \rightarrow \infty} \limsup_{n \rightarrow \infty} \max_{\eta \in \{0,1\}^{S^n}} \|\mu_{\frac{1}{2} \log |S^n| + \alpha}^{n, \eta} - \mu_\infty^n\|_{\text{TV}} = 0.$$

**Remark.** We will see that (1.4) holds in fact whenever  $\lim_{n \rightarrow \infty} |S^n| = \infty$  and therefore the upper bound (1.2) also holds already under this assumption.

Theorem 1 tells us that, under the given conditions, the mixing time is of order  $\frac{1}{2} \log |S^n|$  and that there is a cutoff with a window of size of order 1. (We define mixing times and cutoff in Section 2 below.) These assumptions are necessary. Clearly if there is no bound on  $(q_{\text{max}}^n)$ , then the mixing time

can easily be made to be of order 1. More interestingly, even if  $(q_{\max}^n)$  is bounded, (1.3) is not necessarily true without some condition on the set of stationary distributions. An example of this is continuous time random walk with a small parameter  $c$  on the  $n$ -star which is the graph which has one vertex with  $n$  edges emanating from it. (By continuous time random walk with parameter  $c$  on a graph, we mean that the walker waits an exponential time with parameter  $c$  and then chooses a neighbor at random.) This will be explained in Section 4. We also mention that it is easy to see that the condition involving the set of stationary distributions is not necessary in order for (1.3) to hold since one could take  $(q_{\max}^n)$  going to 0 sufficiently quickly so that the voter mechanism never comes into play.

We mention that it was proved in [12] by Ramadas that when randomization occurs at any rate  $\delta$ , the mixing time for the noisy voter model on any graph with  $n$  vertices is  $O_\delta(\log n)$ .

Theorem 1 has an interesting consequence for the stochastic Ising model on cycles. The Ising model on any graph  $G = (V, E)$  with parameter (inverse temperature)  $\beta \geq 0$  is the probability measure on  $\{-1, 1\}^V$  which gives, up to a normalization factor, probability  $e^{\beta \sum_{\{x,y\} \in E} \sigma(x)\sigma(y)}$  to configuration  $\sigma$ . The stochastic Ising model on  $G$  with *heat-bath* dynamics is the continuous time Markov chain on  $\{-1, 1\}^V$  where each site at rate 1, erases its present state and chooses to be in state 1 or 0 according to the conditional distribution for the Ising model given the other states at that time. For the case  $(\mathbb{Z}/n\mathbb{Z})^d$ , Lubetzky and Sly (see [7]) proved that for  $d = 1$  and all  $\beta$ ,  $d = 2$  and all  $\beta$  below the critical value and  $d = 3$  and all  $\beta$  sufficiently small, one has cutoff at some constant times  $\log n$  with a window of order  $\log \log n$ . In [8], Lubetzky and Sly improved and extended these results in a number of directions; in particular they proved that the result holds for all  $\beta$  below the critical value in all dimensions and that the window above can be taken to be of order 1. While the arguments in this second paper are somehow easier, they are still quite involved, including for  $d = 1$ .

Interestingly, for the cycle  $\mathbb{Z}/n\mathbb{Z}$ , the stochastic Ising model and the noisy voter model (where one performs random walk on  $\mathbb{Z}/n\mathbb{Z}$ ) turn out to be the same model and hence the special case of Theorem 1 for random walk on the cycle was already known. In this special case, the stochastic Ising model corresponds to the dynamics where independently at each  $x \in S^n$ , the rate at which  $\sigma(x)$  flips to  $-\sigma(x)$  is

$$(1.5) \quad \left[ 1 + \exp \left( 2\beta\sigma(x)[\sigma(x-1) + \sigma(x+1)] \right) \right]^{-1}.$$

An easy calculation left to the reader shows that if we consider the noisy voter model on the cycle with  $q(x, x+1) = q(x, x-1) = (e^{4\beta} - 1)/4$  and multiply time by  $\theta := \frac{2}{1+e^{4\beta}}$ , we obtain the above stochastic Ising model. While the work of Lubetzky and Sly implies Theorem 1 for the cycle (and also yields some further results), the proof given here turns out to be easier.

Mossel and Schoenebeck ([9]) consider a similar type of voting model where there is no noise and study, among other things, the time it takes to become absorbed. Here, properly related to our model, they show an upper bound of order  $n^3$  which would be the correct order for the cycle. We see from the last part of Theorem 1 a drastic change when even small noise is introduced into the system since now it takes only order  $n \log n$  to reach equilibrium. On a related note, Mossel and Tamuz ([10]) provide a fascinating survey of various “opinion exchange dynamics”.

We earlier mentioned the  $n$ -star as providing a counterexample to Theorem 1.3 when there is no condition imposed on the stationary distributions. The noisy voter model on the  $n$ -star has an additional fascinating feature.

**Theorem 2.** *Consider the noisy voter model corresponding to continuous time random walk with parameter 1 on the  $n$ -star with  $n$  even.*

(i). *Let  $\eta_0$  denote any configuration which is 1 on exactly half of the leaves. If  $n \geq 3$  and  $t = \frac{1}{4}(\log n - C) > 0$  then*

$$(1.6) \quad \|\mu_t^{\eta_0} - \mu_\infty\|_{\text{TV}} \geq \frac{e^C}{48 + e^C}.$$

(ii). *The time it takes for the distribution starting from all 1’s to be within distance  $1/4$  in total variation norm from the stationary distribution is  $O(1)$ .*

This is quite surprising since one typically expects that for monotone systems, the mixing time for the system should be governed by the time it takes the two extremal states to become close in total variation norm.

We end this introduction with a brief description of the results obtainable for a natural version of a discrete time noisy voter model. The input for such a model is a discrete time Markov chain on a finite set  $S$  and a parameter  $\gamma \in [0, 1]$ . Given these, the model is defined by first choosing an  $x$  in  $S$  uniformly at random, and then with probability  $1 - \gamma$ , one selects  $y$  with probability  $P(x, y)$  at which point the state of  $x$  changes to the state of  $y$  while with probability  $\gamma$ , the state at vertex  $x$  is rerandomized to be 0 or 1, each with probability  $1/2$ . Discrete time analogues of (1.3) (and (3.1) later on) can easily be obtained with the exact same methods to be used below. The mixing times however will be at time  $\frac{|S| \log |S|}{2\gamma}$  now since we are only updating 1 vertex at a time and rerandomizing with probability  $\gamma$ . Similarly, a discrete time analogue of (1.6) can be obtained when for example  $\gamma = 1/2$ ; here the relevant time will be  $n \log n/2$ . The connection with the Ising model holds exactly when moving to discrete time but then one must consider the discrete time version of the Ising model. The paper by Chen and Saloff-Coste (see [1]) contains various results which allow one to transfer between a discrete time model and its continuous time version (where updates are done at the times of a Poisson process). In particular, Proposition 3.2(2) in this paper allows us to obtain a discrete time analogue of (1.4) (with time scaled again by  $n/\gamma$ ) from the continuous time version of this result. Finally a discrete time analogue of Theorem 2(ii) with the

$O(1)$  term being replaced by an  $O(n)$  term can be obtained; this is done by modifying the proof of Theorem 20.3(ii) in [6] to obtain a discrete time version of Lemma 1 from the continuous time version of this lemma.

The rest of the paper is organized as follows. In Section 2, we briefly recall some standard definitions concerning mixing times and cutoff as well as introduce some notation. In Section 3 we prove a stronger version of Theorem 1, namely Theorem 3. The coalescing Markov chain description of both the voter model and the noisy voter model is an important tool in its analysis. However, in this paper, we only need this tool for the proof of the last statement of Theorem 1 or equivalently for Theorem 3(ii) (as well as in the first remark in Section 4) and therefore this description is discussed only at that point in the paper. Lastly, Theorem 2 is proved in Section 4.

## 2. BACKGROUND

In this section, we recall some standard definitions. Consider a continuous time irreducible Markov chain on a finite set  $\Omega$  with transition matrices  $\{P^t(x, y)\}_{t \geq 0}$  and stationary distribution  $\pi$ . Letting  $P^t(x, \cdot)$  denote the distribution at time  $t$  starting from  $x$ , we let

$$(2.1) \quad d(t)(x) := \|P^t(x, \cdot) - \pi\|_{\text{TV}}, \quad \bar{d}(t)(x, y) := \|P^t(x, \cdot) - P^t(y, \cdot)\|_{\text{TV}}$$

and

$$d(t) := \max_{x \in \Omega} d(t)(x) \quad \bar{d}(t) := \max_{x, y \in \Omega} \bar{d}(t)(x, y).$$

Next for  $\varepsilon > 0$ , we let  $t_{\text{mix}}(\varepsilon) := \inf\{t \geq 0 : d(t) \leq \varepsilon\}$  denote the  $\varepsilon$ -mixing time and then one takes by convention  $t_{\text{mix}} := t_{\text{mix}}(1/4)$  and calls this the *mixing time*.

The following notions are very natural but are perhaps not standard. For  $\varepsilon > 0$ , we also let  $t_{\text{mix}}(\varepsilon)(x) := \inf\{t \geq 0 : d(t)(x) \leq \varepsilon\}$  and  $t_{\text{mix}}(x) := t_{\text{mix}}(1/4)(x)$ .

Following [6], we say that a sequence of Markov chains exhibits *cutoff* if for all  $\varepsilon > 0$ , we have

$$\lim_{n \rightarrow \infty} \frac{t_{\text{mix}}^n(\varepsilon)}{t_{\text{mix}}^n(1 - \varepsilon)} = 1.$$

We say that a sequence of Markov chains exhibits *cutoff* with a window of size  $w_n$  if  $w_n = o(t_{\text{mix}}^n)$  and in addition

$$\lim_{\alpha \rightarrow \infty} \liminf_{n \rightarrow \infty} d_n(t_{\text{mix}}^n - \alpha w_n) = 1 \quad \text{and} \quad \lim_{\alpha \rightarrow \infty} \limsup_{n \rightarrow \infty} d_n(t_{\text{mix}}^n + \alpha w_n) = 0.$$

For continuous time random walk with rate 1 on the hypercube of dimension  $n$ , it is known (see [3]) that  $t_{\text{mix}}^n \sim \frac{1}{4} \log n$  and that there is cutoff with a window of order 1. Theorem 1 states that for the noisy voter model, under the given assumptions, we have that  $t_{\text{mix}}^n \sim \frac{1}{2} \log n$  and that there is cutoff with a window of order 1. (The difference of  $\frac{1}{4}$  and  $\frac{1}{2}$  here is simply due to the fact that continuous time random walk with rate 1 on the hypercube of dimension  $n$  has each coordinate changing its state at rate 1 rather than rerandomizing at rate 1.) We point out that in most cases where cutoff is

proved, the chain is reversible while Theorem 1 provides for us a large class of nonreversible chains.

### 3. PROOF OF THEOREM 1

We state here a stronger and more detailed version of Theorem 1. First, given any probability measure on a set, we let

$$\pi_{\max} := \max_{x \in S} \pi(x), \quad \pi_{\min} := \min_{x \in S} \pi(x) \quad \text{and} \quad \rho(\pi) := \frac{\pi_{\max}}{\pi_{\min}}.$$

Given  $S$  and  $q$  as above, we let  $\mathcal{D}(q)$  denote the collection of stationary distributions and let

$$\rho(q) := \min_{\pi \in \mathcal{D}(q)} \rho(\pi).$$

**Theorem 3.** (i). *Fix  $S$  and  $q$ . Let  $\mathbf{1}$  denote the configuration of all 1's and  $\alpha \geq 1$  and assume that  $t := \frac{1}{2} \log |S| - \alpha \geq 1$ . Then*

$$(3.1) \quad \|\mu_t^{\mathbf{1}} - \mu_\infty\|_{\text{TV}} \geq \frac{.7e^{2\alpha}}{16(1 + q_{\max})^2 \rho^2(q) + .7e^{2\alpha}}.$$

(ii). *Fix  $S$  and  $q$ . Letting superscript  $H$  denote random walk (sped down by a factor of 2) on  $\{0, 1\}^S$  (i.e.,  $q \equiv 0$ ), we have that for all  $t$*

$$(3.2) \quad \max_{\eta_1, \eta_2 \in \{0, 1\}^S} \|\mu_t^{\eta_1} - \mu_t^{\eta_2}\|_{\text{TV}} \leq \max_{\eta_1, \eta_2 \in \{0, 1\}^S} \|\mu_t^{\eta_1, H} - \mu_t^{\eta_2, H}\|_{\text{TV}}.$$

One observes that (3.1) implies (1.3) under the assumptions given in Theorem 1. Next, since  $\max_{\eta_1, \eta_2 \in \{0, 1\}^S} \|\mu_{\frac{1}{2} \log |S| + \alpha}^{\eta_1, H} - \mu_{\frac{1}{2} \log |S| + \alpha}^{\eta_2, H}\|_{\text{TV}}$  is (see [3]) at most  $\frac{4}{\sqrt{\pi}} \int_0^{\frac{e^{-\alpha}}{\sqrt{8}}} e^{-t^2} dt + o(1)$  as  $|S| \rightarrow \infty$ , we have that (3.2) implies (1.4) under the assumption that  $\lim_{n \rightarrow \infty} |S^n| = \infty$ .

#### 3.1. Proof of Theorem 3(i).

*Proof of Theorem 3(i).* We will apply Wilson's method for obtaining lower bounds on mixing times; see [13] or Section 13.2 in [6]. Choose  $\pi \in \mathcal{D}(q)$  which minimizes  $\rho(\pi)$  and let  $\Phi(\eta) := 2 \sum_{x \in S} \eta(x) \pi(x) - 1$ . We claim that we have that

$$(3.3) \quad \mathbb{E}_\eta[\Phi(\eta_t)] = e^{-t} \Phi(\eta).$$

To see this, let  $\eta^x$  denote the configuration  $\eta$  except that the coordinate at  $x$  is changed to  $1 - \eta(x)$ , and note that  $\Phi(\eta^x) - \Phi(\eta) = 2\pi(x)(1 - 2\eta(x))$ .

Then by (1.1),

$$\begin{aligned} & \frac{d}{dt} \mathbb{E}_\eta(\Phi(\eta_t))|_{t=0} \\ &= \sum_{x \in S} \left( \frac{1}{2} + \sum_{y \neq x} q(x, y) 1\{\eta(y) \neq \eta(x)\} \right) 2\pi(x)(1 - 2\eta(x)) \\ &= -\Phi(\eta) + 2 \sum_{x, y \neq x} \pi(x) q(x, y) 1\{\eta(y) \neq \eta(x)\} (1 - 2\eta(x)). \end{aligned}$$

A calculation using the stationarity of  $\pi$  shows that the last sum is zero. This proves  $\frac{d}{dt} \mathbb{E}_\eta(\Phi(\eta_t))|_{t=0} = -\Phi(\eta)$ , and hence (3.3) holds.

Next we claim that for any  $t$

$$(3.4) \quad \mathbb{E}_\eta (|\Phi(\eta_t) - \Phi(\eta)|^2) \leq (2\pi_{\max})^2 [ |S|(1 + q_{\max})t + (|S|(1 + q_{\max})t)^2 ].$$

This is because a jump of  $\eta_t$  changes  $\Phi$  by at most  $2\pi_{\max}$ , while by (1.1) the number of jumps during the interval  $[0, t]$  is stochastically dominated above by a Poisson random variable with mean  $|S|(1 + q_{\max})t$ .

Now consider the *discrete* time Markov chain obtained by sampling  $\eta_t$  at times which are integer multiples of  $1/|S|$ . Then  $\Phi$  is an eigenfunction for this discrete time chain with eigenvalue  $\lambda := e^{-1/|S|} \in (\frac{1}{2}, 1)$  (if  $|S| \geq 2$ ). We can now apply equation (13.9) from Section 13.2 of [6] to this discrete time Markov chain with  $t$  being  $|S|(\frac{1}{2} \log |S| - \alpha)$ ,  $x$  being the configuration  $\mathbf{1}$  (whose corresponding  $\Phi$  value is 1) and  $R$  being  $8\pi_{\max}^2(1 + q_{\max})^2$  (see (3.4)). Using  $\pi_{\max} \leq \rho(q)/|S|$  and multiplying the numerator and denominator of the obtained fraction from (13.9) in [6] by  $|S|^2$  yields (3.1); recall our continuous time system at time  $\frac{1}{2} \log |S| - \alpha$  is the discrete time system at time  $|S|(\frac{1}{2} \log |S| - \alpha)$ .  $\square$

**3.2. Proof of Theorem 3(ii).** For part Theorem 3(ii), we need to recall for the reader the graphical representation for the noisy voter model in terms of coalescing Markov chains. In preparation for this part of the proof, we will also give a result of Evans, Kenyon, Peres and Schulman (see [4]) concerning channels for noisy trees.

We construct our  $(S, q)$  noisy voter model using a so-called graphical representation. Figure 1 illustrates the different elements that arise in the graphical representation. The meaning of the trees, depicted by the dotted, solid and dashed lines will be discussed when we get to the proof of Theorem 3(ii). We start with the random voting times and random choices,  $T^x = \{T_n^x, n \geq 1\}$  and  $W^x = \{W_n^x, n \geq 1\}$ ,  $x \in S$ . The  $T^x$  are independent Poisson processes,  $T^x$  has rate  $q(x) := \sum_{y \neq x} q(x, y)$ , and the  $W_n^x$  are independent  $S$ -valued random variables, independent of the Poisson processes, with  $\mathbb{P}(W_n^x = y) = q(x, y)/q(x)$  for  $x \neq y$ . The rerandomization times and places are given by  $R^x = \{R_n^x, n \geq 1\}$  and  $Z^x = \{Z_n^x, n \geq 1\}$ ,  $x \in S$ . The  $R^x$  are independent rate 1 Poisson processes, and the  $Z_n^x$  are i.i.d. Bernoulli random variables,  $\mathbb{P}(Z_n^x = 1) = \mathbb{P}(Z_n^x = 0) = 1/2$ .

Given  $\eta_0 \in \{0, 1\}^S$ , we define  $\eta_t, t > 0$  as follows. (i) At the times  $t = T_n^x$ , we draw an arrow  $(x, T_n^x) \rightarrow (W_n^x, T_n^x)$  and set  $\eta_t(x) = \eta_{t-}(W_n^x)$ . (ii) At the times  $t = R_n^x$ , we put a  $*$  at  $(x, t)$  and set  $\eta_t(x) = Z_n^x$ . A little thought shows that  $\{\eta_t\}_{t \geq 0}$  has the dynamics specified by (1.1).

We construct the usual voter model dual process of coalescing Markov chains. For  $x \in S$  and  $t > 0$  we construct  $B_s^{x,t}, 0 \leq s \leq t$  as follows. Set  $B_0^{x,t} = x$  and then let  $B_s^{x,t}$  trace out a path going backwards in time to time 0, following the arrows for jumps. More precisely, if  $T^x \cap (0, t) = \emptyset$  put  $B_s^{x,t} = x$  for  $0 \leq s \leq t$ . Otherwise, let  $k = \max\{n \geq 1 : T_n^x < t\}$  and  $u = T_k^x$ , and set

$$B_s^{x,t} = x \text{ for } 0 < s < t - u \text{ and } B_{t-u}^{x,t} = W_k^x.$$

We continue this process starting at  $(B_{t-u}^{x,t}, t - u)$ , thus defining  $B_s^{x,t}$  for all  $0 \leq s \leq t$ . Observe that for each  $x \in S$ ,  $B_s^{x,t}$  is a  $q$ -Markov chain starting at  $x$ . Also, these chains are independent until they meet, at which time they coalesce and move together thereafter.

For  $t > 0$  introduce  $\Pi_t = \{(y, R_k^y), y \in S, k \geq 1 : R_k^y \leq t\}$  which contains all information up to time  $t$  concerning the rerandomization times. For each  $x \in S$ , we want to look at the time it takes the chain  $B_s^{x,t}$  to first encounter a rerandomization event, and also the rerandomization choice. We do this as follows. If  $(B_s^{x,t}, t - s) \notin \Pi_t$  for all  $0 \leq s \leq t$  put  $e(x, t) = \infty$ . Otherwise, let  $y, k$  satisfy  $B_{t-R_k^y}^{x,t} = y$  and  $(B_s^{x,t}, t - s) \notin \Pi_t$  for  $s < t - R_k^y$ , and put  $e(x, t) = t - R_k^y$  and  $Z(x, t) = Z_k^y$ . Given any  $\eta \in \{0, 1\}^S$ , the noisy voter model  $\eta_t^\eta$  with initial state  $\eta_0^\eta = \eta$  can be represented as

$$(3.5) \quad \eta_t^\eta(x) = Z(x, t)1\{e(x, t) \leq t\} + \eta(B_t^{x,t})1\{e(x, t) > t\}$$

and this representation will be assumed in the rest of the proof.

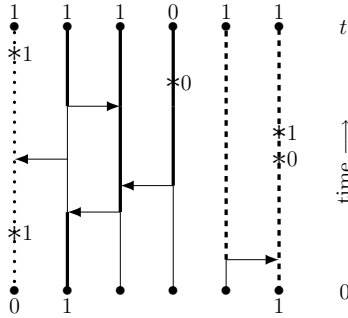


FIGURE 1. The graphical representation and its associated trees: arrows represent voting moves and asterisks represent rerandomization times. In this realization, there are three trees.

In our proof of Theorem 3(ii) we will use the above graphical construction to construct certain *noisy trees* and their associated *stringy trees*. A



noisy tree  $T$  is a tree with flip probabilities in  $(0, \frac{1}{2}]$  labeling the edges. Its associated *stringy tree*  $\widehat{T}$  is the tree which has the same set of root-leaf paths as  $T$  but in which these paths act independently. More precisely, for every root-leaf path in  $T$ , there exists an identical (in terms of length and flip probabilities on the edges) root-leaf path in  $\widehat{T}$ , and in addition, all the root-leaf paths in  $\widehat{T}$  are edge-disjoint. See Figure 2 for an example.

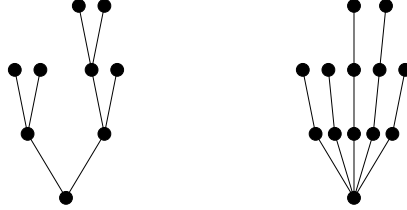


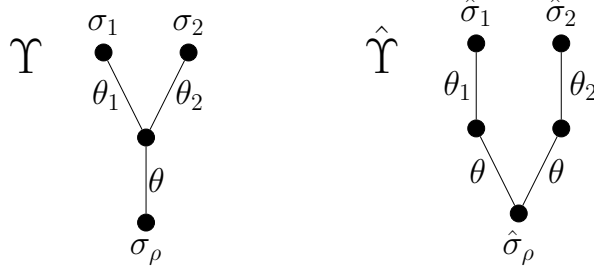
FIGURE 2. A tree  $T$  and the corresponding stringy tree  $\widehat{T}$ .

Starting with  $\sigma_\rho \in \{-1, +1\}$  uniform at the root  $\rho$  of  $T$ , we proceed upward along the tree, assigning a value to each vertex by independently reversing the value of the state of the parent vertex with the probability assigned to the connecting edge (and retaining the value otherwise). Theorem 6.1 in [4] relates the conditional joint distribution (given  $\sigma_\rho$ ) of the resulting variables  $\sigma_w$ , where  $w$  is a leaf of  $T$  with the corresponding conditional joint distribution (given  $\sigma_{\hat{\rho}}$ ) for the associated stringy tree  $\widehat{T}$  using *channels*. If  $X$  is a random variable taking values in  $\Omega_X$  and  $Y$  is a random variable taking values in  $\Omega_Y$ , a *channel* from  $X$  to  $Y$  is a mapping  $f : \Omega_X \times [0, 1] \rightarrow \Omega_Y$  such that if  $Z$  is a uniform random variable on  $[0, 1]$  independent of  $X$ , then  $f(X, Z)$  has distribution  $Y$ . See Section 15.6 in [2].

**Theorem 4** (Theorem 6.1 in [4]). *Given a finite noisy tree  $T$  with leaves  $W$  and root  $\rho$ , let  $\widehat{T}$ , with leaves  $\widehat{W}$  and root  $\hat{\rho}$ , be the stringy tree associated with  $T$ . There is a channel which, for  $\xi \in \{\pm 1\}$ , transforms the conditional distribution  $\sigma_{\widehat{W}} | (\sigma_{\hat{\rho}} = \xi)$  into the conditional distribution  $\sigma_W | (\sigma_\rho = \xi)$ . Equivalently, we say that  $\widehat{T}$  dominates  $T$ .*

**Sketch of Proof.** Our sketch of proof is motivated by and very similar to the proof sketch given in [11]. We only establish a key special case of the theorem: namely, that the tree  $\Upsilon$  shown in Figure 3, is dominated by the corresponding stringy tree  $\widehat{\Upsilon}$ . The general case is derived from it by applying an inductive argument; see [4] for details.

Let  $\theta, \theta_1, \theta_2 \in (0, \frac{1}{2}]$  be the edge flip probabilities in Figure 2, and assume neither of  $\theta_1$  or  $\theta_2$  equal  $\frac{1}{2}$  (otherwise the identity channel will work), and w.l.o.g. assume also that  $\theta_1 \leq \theta_2$ . Let  $\sigma_\rho = \hat{\sigma}_\rho$ , and let  $z$  be a  $\pm 1$ -valued random variable, independent of the edge flip variables, with mean  $(1 -$

FIGURE 3.  $\Upsilon$  is dominated by  $\hat{\Upsilon}$ .

$2\theta_2)/(1 - 2\theta_1) \in (0, 1]$ . Given  $0 \leq \alpha \leq 1$ , to be specified below, we define the channel as follows:

$$(3.6) \quad \sigma_1^* = \hat{\sigma}_1 \quad \text{and} \quad \sigma_2^* = \begin{cases} \hat{\sigma}_2 & \text{with probability } \alpha \\ \hat{\sigma}_1 z & \text{with probability } 1 - \alpha \end{cases}.$$

It suffices to prove, for the appropriate choice of  $\alpha$ , that  $(\sigma_\rho, \sigma_1, \sigma_2)$  and  $(\hat{\sigma}_\rho, \sigma_1^*, \sigma_2^*)$  have the same distribution, and for this it is enough to show that the means of all corresponding products are equal. (This is a special case of the fact that the characters on any finite Abelian group  $G$  form a basis for the vector space of complex functions on  $G$ .) By symmetry it is only the pair correlations which require work.

Let  $\gamma = 1 - 2\theta$  and  $\gamma_i = 1 - 2\theta_i$ ,  $i = 1, 2$ . Clearly  $\mathbb{E}(\hat{\sigma}_\rho \sigma_1^*) = \mathbb{E}(\sigma_\rho \sigma_1)$ ,  $\mathbb{E}(\hat{\sigma}_\rho \hat{\sigma}_1) = \gamma \gamma_1$  and  $\mathbb{E}(\hat{\sigma}_\rho \hat{\sigma}_2) = \gamma \gamma_2$ , whence  $\mathbb{E}(\hat{\sigma}_\rho \sigma_2^*) = \gamma \gamma_2 = \mathbb{E}(\sigma_\rho \sigma_2)$  for any choice of  $\alpha$ . Finally, from  $\mathbb{E}(\hat{\sigma}_1 \hat{\sigma}_2) = \gamma^2 \gamma_1 \gamma_2$ , it follows that

$$\mathbb{E}(\sigma_1^* \sigma_2^*) = \alpha \gamma^2 \gamma_1 \gamma_2 + (1 - \alpha) \frac{\gamma_2}{\gamma_1} = \gamma_1 \gamma_2 \left[ \alpha \gamma^2 + (1 - \alpha) \frac{1}{\gamma_1^2} \right].$$

Since  $\gamma^2 < 1$  and  $1/\gamma_1^2 > 1$ , we can choose  $\alpha \in [0, 1]$  so that  $\mathbb{E}(\sigma_1^* \sigma_2^*) = \gamma_1 \gamma_2 = \mathbb{E}(\sigma_1 \sigma_2)$ ; explicitly,

$$(3.7) \quad \alpha = (1 - \gamma_1^2)/(1 - \gamma^2 \gamma_1^2).$$

This proves that  $\hat{\Upsilon}$  dominates  $\Upsilon$ . □

*Proof of Theorem 3(ii).* Fix  $t > 0$  throughout. Now for  $\eta \in \{0, 1\}^S$ , consider the construction of  $\eta_t^\eta$  given in (3.5). Letting  $\mathcal{Z}(t) = \{B_u^{x,t}, x \in S, u \in [0, t]\}$ , we may write

$$\mu_t^\eta = \int \mu_t^\eta(\cdot | \mathcal{Z}(t)) d\mathbb{P}(\mathcal{Z}(t)).$$

Therefore, to prove (3.2), it suffices to prove the stronger fact that for any  $\eta_1, \eta_2 \in \{0, 1\}^S$  and any realization  $\mathcal{Z}$ ,

$$(3.8) \quad \|\mu_t^{\eta_1}(\cdot | \mathcal{Z}) - \mu_t^{\eta_2}(\cdot | \mathcal{Z})\|_{\text{TV}} \leq \max_{\eta_1, \eta_2 \in \{0, 1\}^S} \|\mu_t^{\eta_1, H} - \mu_t^{\eta_2, H}\|_{\text{TV}}.$$

To proceed, we will first give for any  $\eta \in S$  and realization  $\mathcal{Z}$ , a useful alternative description of  $\mu_t^\eta(\cdot | \mathcal{Z})$ . Clearly  $\mathcal{Z}$  yields a finite number of

disjoint trees  $T_1, T_2, \dots, T_m$  which describe the coalescing picture. (In the realization of Figure 1, there are three trees indicated by the dotted, solid and dashed lines.) Each tree has its root sitting at  $S \times \{0\}$  and its leaves sitting at  $S \times \{t\}$  in the space-time diagram. Let  $x_j$  be the root of  $T_j$  and  $L_j$  be the set of leaves; the  $L_j$ 's are disjoint and their union is (identified with)  $S$ . We also let  $\mathcal{V}_j := \{(y_k, t_k)\}_{k \in L_j}$  consist of the root, the leaves  $L_j$  and all the branch points of  $T_j$  and we view  $\mathcal{V}_j$  as a tree. (If at time  $s$ , a chain moves from  $w$  to  $z$  coalescing with another walker, then we consider the branch point to be  $(z, s)$  rather than  $(w, s)$ .) None of this depends on the configuration  $\eta$ . Note that the branching is always at most 2 and that the tree can move from one vertical line to another; see the solid tree in Figure 1.

Let  $Y^{\eta, j}$  be the process  $\{\eta_s^\eta\}_{s \leq t}$  conditioned on  $\mathcal{Z}$  restricted to  $\mathcal{V}_j$ . (This process also depends of course on  $t$  and  $\mathcal{Z}$  but its dependence on  $\eta$  is what we wish to emphasize.) Next, conditioned on  $\mathcal{Z}$ ,  $Y^{\eta, 1}, Y^{\eta, 2}, \dots, Y^{\eta, m}$  are clearly independent since  $Y^{\eta, j}$  depends only on  $\Pi_t \cap T_j$  and the corresponding  $Z_n^x$ 's. (This implies of course that  $\eta_t^\eta(L_1), \eta_t^\eta(L_2), \dots, \eta_t^\eta(L_m)$  are conditionally independent given  $\mathcal{Z}$ .) We also let  $Y^\eta$  be the process  $\{\eta_s^\eta\}_{s \leq t}$  restricted to  $\cup_j \mathcal{V}_j$ . Crucially,  $Y^{\eta, j}$  has the following alternative simpler description as a tree-indexed Markov chain which is easy to verify and left to the reader.

At the root  $(x_j, 0)$  of  $\mathcal{V}_j$ , the value of  $Y^{\eta, j}$  is  $\eta(x_j)$ . Inductively, the value of  $Y^{\eta, j}$  at a particular node is taken to be the same as the value of its parent node (which is lower down on the time axis) with probability  $\frac{1+e^{-s}}{2}$  where  $s$  is the time difference between these two nodes, and the opposite value otherwise. These random choices are taken independently. The dependence of  $Y^{\eta, j}$  on  $\eta$  is only through the initial state  $\eta(x_j)$ ; otherwise, the transition mechanism is the same.

Consider now the process  $\tilde{Y}^\eta$  indexed by  $S$  defined by the following two properties: the random variables  $\tilde{Y}^\eta(x), x \in S$  are independent, and for each  $j$ , for all  $x \in L_j$ ,  $\tilde{Y}^\eta(x) = \eta(x_j)$  with probability  $\frac{1+e^{-t}}{2}$  and the opposite value otherwise. It is easy to see that the distribution of  $\tilde{Y}$  is simply the distribution for continuous time random walk on the hypercube at time  $t$  started from the configuration whose state at  $x$  is  $\eta(x_j)$  for  $x \in L_j, j = 1, \dots, m$ .

Theorem 4 now implies that for each  $j$ , there is a channel (depending on  $T_j$ ) *not depending on*  $\eta(x_j)$  which transforms the random variables  $\tilde{Y}^\eta(L_j)$  to the random variables  $Y^\eta(L_j) = Y^{\eta, j}(L_j)$  meaning that given the tree  $T_j$ , there is a function

$$f_j : \{0, 1\}^{L_j} \times [0, 1] \rightarrow \{0, 1\}^{L_j}$$

so that if  $U$  is a uniform random variable on  $[0, 1]$  independent of everything else, we have that for each value of  $\eta(x_j)$

$$f_j(\tilde{Y}^\eta(L_j), U) \text{ and } Y^\eta(L_j)$$

are equal in distribution.

Since  $\tilde{Y}^\eta(L_j)$  are independent as we vary  $j$  and similarly for  $Y^\eta(L_j)$ , it follows that we have a function (depending on  $\mathcal{Z}$ )

$$f : \{0, 1\}^S \times [0, 1] \rightarrow \{0, 1\}^S$$

so that if  $U$  is a uniform random variable on  $[0, 1]$  independent of everything else, we have that for any  $\eta$

$$f(\tilde{Y}^\eta(S), U) \text{ and } Y^\eta(S)$$

are equal in distribution.

This then easily yields that for any  $\eta_1$  and  $\eta_2$ ,

$$\|Y^{\eta_1}(S) - Y^{\eta_2}(S)\|_{\text{TV}} \leq \|\tilde{Y}^{\eta_1}(S) - \tilde{Y}^{\eta_2}(S)\|_{\text{TV}}.$$

Finally, it is clear from construction that

$$\|\tilde{Y}^{\eta_1}(S) - \tilde{Y}^{\eta_2}(S)\|_{\text{TV}} \leq \max_{\eta_1, \eta_2 \in \{0, 1\}^S} \|\mu_t^{\eta_1, H} - \mu_t^{\eta_2, H}\|_{\text{TV}},$$

completing the proof.  $\square$

#### 4. THE $n$ -STAR AND THE PROOF OF THEOREM 2

In this section, we consider the noisy voter model  $\{\eta_t^n\}$  on the the  $n$ -star. We first explain why this gives us an example showing that the conclusion of Theorem 1(i) is not true in general without the assumption of a uniform bound on the  $\rho_n$ 's even if  $(q_{\max}^n)$  is bounded. Consider first continuous time random walk on the  $n$ -star with rate 1, meaning that the walker waits an exponential amount of time with parameter 1 and then moves to a uniform neighbor. If we run a corresponding system of coalescing Markov chains starting from each point, it is not hard to see that the expected time until all chains coalesce is at most  $O(\log n)$ . If we now multiply all the rates by a certain large constant  $c$ , we will have that the expected time until all chains coalesce is at most  $\log n/32$ . Then by Markov's inequality, the probability that the chains have not coalesced by time  $\log n/4$  is at most  $1/8$ . Since each site is rerandomized at rate 1, it is easy to see from this fact and the graphical construction in Section 3 that this implies that the mixing time at most  $\log n/3$ .

The rest of the section is devoted to the proof of Theorem 2.

*Proof of Theorem 2.* We begin with (i). This is similar to the proof of Theorem 3(i), except one considers a different eigenfunction. Partition the leaves into disjoint sets  $A$  and  $B$  each with  $n/2$  elements. Let

$$\Phi(\eta) := \sum_{x \in A} \eta(x) - \sum_{x \in B} \eta(x).$$

It is elementary to check that

$$(4.1) \quad \mathbb{E}_\eta[\Phi(\eta_t)] = e^{-2t}\Phi(\eta).$$

As in the proof of Theorem 3(i), we consider the time Markov chain obtained by sampling our process at times which are integer multiples of

$1/n$ . Then  $\Phi$  is an eigenfunction for this discrete time chain with eigenvalue  $\lambda := e^{-2/n} \in (\frac{1}{2}, 1)$  (if  $n \geq 3$ ). We can now apply equation (13.9) from Section 13.2 of [6] to this discrete time Markov chain with  $t$  being  $\frac{n}{4}(\log n - C)$ ,  $x$  being the configuration  $\eta_0$  (whose corresponding  $\Phi$  value is  $n/2$ ) and  $R$  being 6. After simplification (and recall our continuous time system at time  $\frac{1}{4}(\log n - C)$  is the discrete time system at time  $\frac{n}{4}(\log n - C)$ ) yields (1.6).

For (ii), note first that, in the terminology introduced in Section 2, we want to show that  $t_{\text{mix}}^n(\mathbf{1}) = O(1)$ . We first note that by symmetry, if we only look at the state of the center of the star and the number of leaves which are in state 1, then this is also a Markov chain. (It is a *projection* of the original chain in the sense of Section 2.3.1 in [6].) Let  ${}^R\eta_t^n$  denote this “reduced” Markov chain whose state space is  $\{0, 1\} \times \{0, 1, \dots, n\}$ . The key step in proving that  $t_{\text{mix}}^n(\mathbf{1}) = O(1)$  is to show that this reduced chain has mixing time  $O(1)$  which is interesting in itself; this is stated in Lemma 1 below.

Assuming this lemma, one proceeds as follows. By symmetry considerations, we can generate a realization of the configuration at time  $t$  starting from all 1’s by considering the reduced system at time  $t$  starting from  $(1, n)$  and if the reduced system is in state  $(a, k)$ , we then construct a configuration for the full system by letting the center be in state  $a$  and choosing a uniform random subset of size  $k$  from the  $n$  leaves to be in state 1 and the rest to be in state 0. We can generate a realization from the stationary distribution for the full system in an analogous way by choosing  $(a, k)$  from the stationary distribution of the reduced system and then letting the center be in state  $a$  and choosing a uniform random subset of size  $k$  from the  $n$  leaves to be in state 1 and the rest to be in state 0. Therefore, by an obvious coupling, we have that the total variation distance between the full system at time  $t$  started from  $\mathbf{1}$  and the stationary distribution for the full system is exactly the total variation distance between the reduced system at time  $t$  started from  $(1, n)$  and the stationary distribution for the reduced system. Now the proposition follows from Lemma 1.  $\square$

**Lemma 1.** *The mixing times for  $\{{}^R\eta_t^n\}$  is  $O(1)$ .*

*Proof.* Observe that the infinitesimal rates for this reduced chain are as follows.

$$\begin{aligned} (0, k) &\rightarrow (1, k) \text{ at rate } \frac{1}{2} + \frac{k}{n} \\ (0, k) &\rightarrow (0, k+1) \text{ at rate } \frac{n-k}{2} \\ (0, k) &\rightarrow (0, k-1) \text{ at rate } \frac{3k}{2} \\ (1, k) &\rightarrow (0, k) \text{ at rate } \frac{1}{2} + \frac{n-k}{n} \\ (1, k) &\rightarrow (1, k+1) \text{ at rate } \frac{3(n-k)}{2} \end{aligned}$$

$$(1, k) \rightarrow (1, k - 1) \text{ at rate } \frac{k}{2}$$

We denote this reduced system by  $(X_t, Y_t)$  where  $n$  is suppressed in the notation. The key fact that we will show is that there exists  $c_1 > 0$  so that for all  $n$ , for all (initial) states  $(a, \ell)$ , for all (final) states  $(b, k)$  with  $k \in [4n, .6n]$ ,

$$\mathbb{P}_{(a, \ell)}((X_{10}, Y_{10}) = (b, k)) \geq c_1/n.$$

By equation (4.13) in [6], this implies that there exists  $c_2 > 0$  so that for all  $n$ , for any two initial states, the total variation distance of the corresponding processes at time 10 is at most  $1 - c_2$ . This easily leads to the claim of the lemma.

Since it is very easy for the center to change states, it is easy to see that it suffices to prove the above key fact when  $a = 1$  and  $b = 0$ .

Let  $U$  be the event that the center never during  $[0, 10]$  attempts an update by looking at one of its neighbors. Letting  $A_t := U \cap \{X_s = 1 \forall s \in [0, t]\}$ , one checks that the conditional distribution of  $Y_t$  given  $A_t$  is the sum of two independent binomial distributions with respective parameters  $(\ell, \frac{3}{4} + \frac{1}{4}e^{-2t})$  and  $(n - \ell, \frac{3}{4} - \frac{3}{4}e^{-2t})$ . In particular

$$g(t) := \mathbb{E}\left[\frac{Y_t}{n} | A_t\right] = \frac{3}{4} + \left(\frac{\ell}{n} - \frac{3}{4}\right)e^{-2t}.$$

One also easily checks that for all  $n$  and  $\ell$

$$(4.2) \quad |g(t) - g(s)| \leq 2|t - s|.$$

The representation of  $Y_t$  as a sum of two binomials when conditioned on  $A_t$  yields  $\text{Var}\left(\frac{Y_t}{n} | A_t\right) \leq 1/n$  and hence by Chebyshev's inequality we have that for all  $n, \ell, t$  and  $\sigma$

$$(4.3) \quad \mathbb{P}_{(a, \ell)}\left(|\frac{Y_t}{n} - g(t)| \geq \frac{\sigma}{\sqrt{n}} | A_t\right) \leq 1/\sigma^2.$$

Now, letting  $B_t := U \cap \{X_s = 0 \forall s \in [t, 10]\}$ , one checks that the conditional distribution of  $Y_{10}$  given  $B_t \cap \{Y_t = nu\}$  is the sum of two independent binomial distributions with respective parameters  $(nu, \frac{1}{4} + \frac{3}{4}e^{-2(10-t)})$  and  $(n(1 - u), \frac{1}{4} - \frac{1}{4}e^{-2(10-t)})$ . In particular

$$h(u, t) := \mathbb{E}\left[\frac{Y_{10}}{n} | B_t \cap \{Y_t = nu\}\right] = \frac{1}{4} + \left(u - \frac{1}{4}\right)e^{-2(10-t)}.$$

One also easily checks that for all  $u, v$  and  $t, s \in [0, 10]$

$$(4.4) \quad |h(u, t) - h(v, s)| \leq 2(|u - v| + |t - s|).$$

By an easy variant of the local central limit theorem, there exists  $c_3 > 0$  so that for all  $n, u, t \in [0, 9.9]$  and integer  $v \in [nh(u, t) - 10\sqrt{n}, nh(u, t) + 10\sqrt{n}]$  one has that

$$(4.5) \quad \mathbb{P}[Y_{10} = v | B_t \cap \{Y_t = nu\}] \geq \frac{c_3}{\sqrt{n}}.$$

Next, one easily checks that  $h(g(0), 0) \leq .4$  and  $h(g(9.9), 9.9) \geq .6$  and hence by our assumptions on  $k$ , there exists  $t^* \in [0, 9.9]$  such that  $h(g(t^*), t^*) = \frac{k}{n}$ . (It is easily checked that  $h(g(t), t)$  is increasing in  $t$  but that is not needed to conclude the existence of  $t^*$ .)

We now let  $G$  be the intersection of the events  $U$  and that during  $[0, 10]$ , the center flips its state exactly once and that this occurs during  $[t^* - \frac{1}{n^{1/2}}, t^* + \frac{1}{n^{1/2}}]$ . Clearly there exists  $c_4 > 0$  so that for all  $n$  and  $t^*$ , we have that  $\mathbb{P}(G) \geq \frac{c_4}{\sqrt{n}}$ . On the event  $G$ , we let  $T$  denote this unique flipping time of the center.

Now, by (4.2),  $|g(T) - g(t^*)| \leq 2/\sqrt{n}$  and hence

$$\{|\frac{Y_T}{n} - g(t^*)| \geq 4/\sqrt{n}\} \subseteq \{|\frac{Y_T}{n} - g(T)| \geq 2/\sqrt{n}\}.$$

Applying (4.3), this yields

$$\mathbb{P}_{(a,\ell)}(|\frac{Y_T}{n} - g(t^*)| \geq \frac{4}{\sqrt{n}} | G, T) \leq 1/4.$$

We therefore have

$$\mathbb{P}_{(a,\ell)}(G \cap H) \geq \frac{c_4}{2\sqrt{n}}$$

where  $H := \{|\frac{Y_T}{n} - g(t^*)| \leq \frac{4}{\sqrt{n}}\}$ . Given this lower bound, to prove the key claim now, it would suffice to show that for all parameters,

$$(4.6) \quad \mathbb{P}_{(a,\ell)}(Y_{10} = k | G \cap H) \geq \frac{c_3}{\sqrt{n}}$$

where  $c_3$  comes from (4.5).

By (4.4),  $|T - t^*| \leq \frac{1}{\sqrt{n}}$  and  $|\frac{Y_T}{n} - g(t^*)| \leq \frac{4}{\sqrt{n}}$  imply that

$$|h(\frac{Y_T}{n}, T) - h(g(t^*), t^*)| \leq \frac{10}{\sqrt{n}}$$

and hence by the definition of  $t^*$ , we have  $|h(\frac{Y_T}{n}, T) - \frac{k}{n}| \leq \frac{10}{\sqrt{n}}$ . Finally (4.6) now follows from (4.5) by conditioning on the exact values of  $T$  and  $Y_T$ , completing the proof.  $\square$

**Remark.** In view of the proof of Theorem 2(ii), it also follows that for the reduced system  $t_{\text{mix}}^n(\varepsilon)(\mathbf{1}) = O(\log(1/\varepsilon))$ .

#### ACKNOWLEDGEMENTS

This work was initiated when the third author was visiting Microsoft Research at Redmond, WA and was continued when the first author was visiting Chalmers University of Technology. JTC acknowledges the support of the Simons Foundation and the National Science Foundation. JES acknowledges the support of the Swedish Research Council and the Knut and Alice Wallenberg Foundation.

## REFERENCES

- [1] Chen, G.-Y. and Saloff-Coste, L., Comparison of cutoffs between lazy walks and Markovian semigroups. *J. Appl. Probab.* **50**, 2013, 943–959.
- [2] Cover T. M. and Thomas J. A., Elements of Information Theory. Second edition. Wiley, New York, 2006.
- [3] Diaconis, P., Graham, R. L. and Morrison, J. A., Asymptotic analysis of a random walk on a hypercube with many dimensions. *Random Structures Algorithms* **1**, 1990, 51-72.
- [4] Evans, W., Kenyon, C., Peres, Y. and Schulman, L. J., Broadcasting on trees and the Ising model. *Ann. Appl. Probab.* **10**, 2000, 410-433.
- [5] Granovsky, B. Madras, N., The noisy voter model. *Stochastic Process. Appl.* **55**, 1995, 2343.
- [6] Levin, D. A., Peres, Y. and Wilmer, E. L., Markov chains and mixing times. *American Mathematical Society*, Providence, RI, 2009.
- [7] Lubetzky, E. and Sly, A., Cutoff for the Ising model on the lattice. *Invent. Math.* **191**, 2013, 719-755.
- [8] Lubetzky, E. and Sly, A., Space-time information percolation for the stochastic Ising model, preprint.
- [9] Mossel, E. and Schoenebeck, G., Reaching consensus on social networks. *Proceedings of 1st Symposium on Innovations in Computer Science*, 2010, 214–229.
- [10] Mossel, E. and Tamuz, O., Opinion exchange dynamics, arXiv:1401.4770, 2014.
- [11] Peres, Y., Probability on trees: an introductory climb. Lectures on probability theory and statistics (Saint-Flour, 1997), 193-280, Lecture Notes in Math., **1717**, Springer, Berlin, 1999.
- [12] Ramadas H., Mixing of the noisy voter model. *Electron. Comm. Probab.* **19**, 2014, no. 17, 1-9.
- [13] Wilson, D., Mixing times of lozenge tiling and card shuffling Markov chains. *Ann. Appl. Probab.* **14**, 2004, 274–325.

DEPARTMENT OF MATHEMATICS, SYRACUSE UNIVERSITY, 215 CARNEGIE BUILDING,  
SYRACUSE, NY 13244-1150, USA

*E-mail address:* [jtcox@syr.edu](mailto:jtcox@syr.edu)

*URL:* <http://as-cascade.syr.edu/profiles/pages/cox-theodore.html>

MICROSOFT RESEARCH, 1 MICROSOFT WAY, REDMOND, WA 98052, USA

*E-mail address:* [peres@microsoft.com](mailto:peres@microsoft.com)

*URL:* <http://research.microsoft.com/~peres/>

DEPARTMENT OF MATHEMATICS, CHALMERS UNIVERSITY OF TECHNOLOGY AND GÖTEBORG  
UNIVERSITY, SE-41296 GOTHENBURG, SWEDEN

*E-mail address:* [steif@math.chalmers.se](mailto:steif@math.chalmers.se)

*URL:* <http://www.math.chalmers.se/~steif/>