Title
Mixing time for the Ising model and random walks

Permalink
https://escholarship.org/uc/item/2b54v885

Author
Ding, Jian

Publication Date
2011

Peer reviewed|Thesis/dissertation

Mixing time for the Ising model and random walks

by

Jian Ding

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Statistics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Yuval Peres, Co-chair
Professor Elchanan Mossel, Co-chair
Professor David Aldous
Professor Fraydoun Rezakhanlou

Spring 2011

Mixing time for the Ising model and random walks

Abstract

Mixing time for the Ising model and random walks

by

Jian Ding

Doctor of Philosophy in Statistics

University of California, Berkeley

Professor Yuval Peres, Co-chair
Professor Elchanan Mossel, Co-chair

In this thesis we study the mixing times of Markov chains, e.g., the rate of convergence of Markov chains to stationary measures. We focus on Glauber dynamics for the (classical) Ising model as well as random walks on random graphs.

We first provide a complete picture for the evolution of the mixing times and spectral gaps for the mean-field Ising model. In particular, we pin down the scaling window, and prove a cutoff phenomenon at high temperatures, as well as confirm the power law at criticality. We then move to the critical Ising model at Bethe lattice (regular trees), where the criticality corresponds to the reconstruction threshold. We establish that the mixing time and the spectral gap are polynomial in the surface area, which is the height of the tree in this special case. Afterwards, we show that the mixing time of Glauber dynamics for the (ferromagnetic) Ising model on an arbitrary n-vertex graph at any temperature has a lower bound of $n \log n/4$, confirming a folklore theorem in the special case of Ising model.

In the second part, we study the random walk on the largest component of the near-supcritical Erdös-Rényi graph. Using a complete characterization of the structure for the near-supercritical random graph, as well as various techniques to bound the mixing times in terms of spectral profile, we obtain the correct order for the mixing time in this regime, which demonstrates a smooth interpolation between the critical and the supercritical regime.

i

Contents

Acknowledgments

This dissertation would not have been possible without the help of many people.

Greatest thanks goes to my amazing advisor Yuval Peres whose help made this dissertation possible. He was a fantastic mentor, who provided a constant stream of problems and offered sharp insights to our questions. He constantly supported me as a GSR such that I could focus on research and visit him at Microsoft research, a truly active place for probability. I enjoyed many times having lunches and dinners with him in various restaurants he brought me to, of which *I love Sushi* and the *Whole foods market* are my favorite. Moreover, countless dinners at his place with his family were just so enjoyable and refreshing.

I am fortunate to work with Eyal Lubetzky, my main collaborator during my first two years of research. We spent so many hours working together in office and at cafes, and amazingly we could indeed sit down together and produce tex file. I improved my English so much by talking with him. Perhaps too much, such that in a trip to Florida I was told that I had a middle-east accent!

Another great experience was working with James Lee, who was a master to challenge me in the just right way. We worked intensively for about half a year on a cover time project joint with Yuval and spent much time thinking of other problems. When we took breaks from the main focus of our potential problems, I enjoyed quite a few introductory lectures by him on discrete geometry and metric embeddings, which initialized my interest in that topic.

I am also grateful to the other members of my dissertation committee, Professors Elchanan Mossel, David Aldous and Fraydoun Rezakhanlou for reading my thesis and also for serving on my qualification exam committee. I enjoyed many useful conversations with Elchanan on mathematical research as well as about life in general. The visit to his place at Weizmann institute was a wonderful experience. It was a pleasure to teach as a GSI for Prof. Ani Adhikari, and I learned a lot from her on how to teach effectively and professionally.

I am indebted to Jennifer Chayes and Christian Borgs for giving me the opportunity to working as an intern at MSR New England for the last summer, and for exposing me to the new field of game theory and social networks.

Special thanks went to my academic siblings. I enjoyed the interactions with Asaf Nachmias, Yun Long, Lionel Levine, Ron Peled, Gabor Pete, and Manju Krishnapur, from whom I learned not only mathematics, but also how to handle the ups and downs in research. I am grateful to Yun for all the help she provided when I first relocated to Seattle: life would have been much more difficult otherwise. I also thank Tonci Antunovic, Subhroshekhar Ghosh, Weiyang Ning, Julia Ruscher, Elisa Celis for the great time we spent together.

I was so fortunate to meet so many wonderful fellow students as well as colleagues, from both our department and Microsoft Research. It was such a pleasure to interact with

Chapter 1

Introduction

The whole thesis deals with the mixing time for Markov chains. In order to describe the mixing-time of the chain (X_t), we require several definitions. For any two distributions ϕ, ψ on Ω, the *total-variation distance* of ϕ and ψ is defined to be

$$\|\phi - \psi\|_{\mathrm{TV}} := \sup_{A \subset \Omega} |\phi(A) - \psi(A)| = \frac{1}{2} \sum_{\sigma \in \Omega} |\phi(\sigma) - \psi(\sigma)| \ .$$

The (worst-case) total-variation distance of (X_t) to stationarity at time t is

$$d_n(t) := \max_{\sigma \in \Omega} \|\mathbf{P}_\sigma(X_t \in \cdot) - \mu_n\|_{\mathrm{TV}} \ ,$$

where $\mathbf{P}_\sigma$ denotes the probability given that $X_0 = \sigma$. The total-variation *mixing-time* of (X_t), denoted by $t_{\mathrm{MIX}}(\varepsilon)$ for $0 < \varepsilon < 1$, is defined to be

$$t_{\mathrm{MIX}}(\varepsilon) := \min \{t : d_n(t) \le \varepsilon\} \ .$$

A related notion is the spectral-gap of the chain, $\mathtt{gap} := 1 - \lambda$, where λ is the largest absolute-value of all nontrivial eigenvalues of the transition kernel.

Consider an infinite family of chains $(X_t^{(n)})$, each with its corresponding worst-distance from stationarity $d_n(t)$, its mixing-times $t_{\mathrm{MIX}}^{(n)}$, etc. We say that $(X_t^{(n)})$ exhibits *cutoff* iff for some sequence $w_n = o\big(t_{\mathrm{MIX}}^{(n)}(\frac{1}{4})\big)$ we have the following: for any $0 < \varepsilon < 1$ there exists some $c_\varepsilon > 0$, such that

$$t_{\mathrm{MIX}}^{(n)}(\varepsilon) - t_{\mathrm{MIX}}^{(n)}(1 - \varepsilon) \le c_\varepsilon w_n \quad \text{for all } n \ . \tag{1.0.1}$$

That is, there is a sharp transition in the convergence of the given chains to equilibrium at time $(1 + o(1))t_{\mathrm{MIX}}^{(n)}(\frac{1}{4})$. In this case, the sequence w_n is called a *cutoff window*, and the

sequence $t_{\text{MIX}}^{(n)}(\frac{1}{4})$ is called a *cutoff point.*

We mainly study two kinds of Markov chains in this thesis: Glauber dynamics for the Ising model and random walks on graphs. We have a short discussion on both topics in what follows.

1.1 Glauber dynamics for the Ising model

The (ferromagnetic)*Ising Model* on a finite graph $G = (V, E)$ with parameter $\beta \geq 0$ and no external magnetic field is defined as follows. Its set of possible *configurations* is $\Omega = \{1, -1\}^V$, where each configuration $\sigma \in \Omega$ assigns positive or negatives *spins* to the vertices of the graph. The probability that the system is at a given configuration σ is given by the *Gibbs distribution* The *Ising Model* on a finite graph $G = (V, E)$ at inverse temperature β with interaction strengths $J = \{J_{uv} \geq 0 : uv \in E\}$ and external field $H = \{H_u : u \in V\}$, is a probability measure μ_G on the configuration space $\Omega = \{\pm 1\}^V$, defined as follows. For each $\sigma \in \Omega$,

$$\mu_G(\sigma) = \frac{1}{Z(J, \beta, H)} \exp\left(\beta \sum_{uv \in E} J_{uv}\sigma(u)\sigma(v) + \sum_u H_u\sigma_u\right), \qquad (1.1.1)$$

where $Z(J, \beta, H)$ is a normalizing constant called the *partition function.* The measure μ_G is also called the *Gibbs measure.* Usually, we consider the case where $J \equiv 1$ and $H \equiv 0$ (i.e., with no external field). When there is no ambiguity regarding the base graph, we sometimes write μ for μ_G.

The *heat-bath Glauber dynamics* for the distribution μ_n is the following Markov Chain, denoted by (X_t). Its state space is Ω, and at each step, a vertex $x \in V$ is chosen uniformly at random, and its spin is updated as follows. The new spin of x is randomly chosen according to μ_n conditioned on the spins of all the other vertices. It can easily be shown that (X_t) is an aperiodic irreducible chain, which is reversible with respect to the stationary distribution μ_n.

We carefully study the mixing time for Glauber dynamics where the underlying graph are complete graphs and regular trees, and we also establish a general lower bound for the mixing time when the underlying graph is arbitrary. We summarize our main results as follows.

- An exhaustive analysis on the evolution for the mixing time of Glauber dynamics for the *Curie-Weiss* model (i.e., the Ising model with underlying graph being complete), as laid out in Chapter 2. Notably, we pin down the scaling window, demonstrate a cutoff phenomenon, as well as prove a power law at criticality. Chapter 2 is largely based on a joint work with Eyal Lubetzky and Yuval Peres [25].

- We study the Ising model on regular trees in Chapter 3, and we establish the power law for the mixing time as well as the inverse-gap at criticality. Chapter 3 is largely based on a joint work with Eyal Lubetzky and Yuval Peres [24].

- Chapter 4 is devoted to the lower bound on the mixing time in general. We show that for ferromagnetic Ising model on arbitrary graphs, the mixing time is necessarily bounded by $n \log n/4$. Chapter 4 is largely based on a joint work with Yuval Peres [29].

1.2   Random walks on random graphs

There is a rich interplay between geometric properties of a graph and the behavior of a random walk on it (see, e.g., [3]). A particularly important parameter is the mixing time, which measures the rate of convergence to stationarity. In this paper we focus on random walks on the classical Erdős-Rényi random graph $\mathcal{G}(n, p)$.

The geometry of $\mathcal{G}(n, p)$ has been studied extensively since its introduction in 1959 by Erdős and Rényi [35]. A well-known phenomenon exhibited by this model, typical in second-order phase transitions of mean-field models, is the *double jump*: For $p = c/n$ with c fixed, the largest component $\mathcal{C}_1$ has size $O(\log n)$ with high probability (w.h.p.), when $c < 1$, it is w.h.p. linear in n for $c > 1$, and for $c = 1$ its size has order $n^{2/3}$ (the latter was proved by Bollobás [11] and Łuczak [60]). Bollobás discovered that the critical behavior extends throughout $p = (1 \pm \varepsilon)/n$ for $\varepsilon = O(n^{-1/3})$, a regime known as the *critical window*.

Only in recent years were the tools of Markov chain analysis and the understanding of the random graph sufficiently developed to enable estimating mixing times on $\mathcal{C}_1$. Fountoulakis and Reed [38] showed that, in the strictly supercritical regime ($p = c/n$ with fixed $c > 1$), the mixing time of random walk on $\mathcal{C}_1$ w.h.p. has order $\log^2 n$. Their proof exploited fairly simple geometric properties of $\mathcal{G}(n, p)$, while the key to their analysis was a refined bound [37] on the mixing time of a general Markov chain. The same result was obtained independently by Benjamini, Kozma and Wormald [6]. There, the main innovation was a decomposition theorem for the giant component. However, the methods of these two papers do not yield the right order of the mixing time when c is allowed to tend to 1.

Nachmias and Peres [70] proved that throughout the critical window the mixing time on $\mathcal{C}_1$ is of order n. The proof there used branching process arguments, which were effective since the critical $\mathcal{C}_1$ is close to a tree.

It was unclear how to interpolate between these results, and estimate the mixing time as the giant component emerges from the critical window, since the methods used for the supercritical and the critical case were so different. The focus of this paper is primarily

on the emerging supercritical regime, where $p = (1 + \varepsilon)/n$ with $\varepsilon^3 n \to \infty$ and $\varepsilon = o(1)$. In this regime, the largest component is significantly larger than the others, yet its size is still sublinear. Understanding the geometry of $\mathcal{C}_1$ in this regime has been challenging: Indeed, even the asymptotics of its diameter were only recently obtained by Riordan and Wormald [78], as well as in [23].

In Part II, we determine the order of the mixing time throughout the emerging supercritical regime (see Subsection 5.1.3 for a formal definition of mixing time). This part is largely based on joint work with Eyal Lubetzky and Yuval Peres [28].

Theorem 1 (supercritical regime). *Let $\mathcal{C}_1$ be the largest component of $\mathcal{G}(n, p)$ for $p = \frac{1+\varepsilon}{n}$, where $\varepsilon \to 0$ and $\varepsilon^3 n \to \infty$. With high probability, the mixing time of the lazy random walk on $\mathcal{C}_1$ is of order $\varepsilon^{-3} \log^2(\varepsilon^3 n)$.*

While the second largest component $\mathcal{C}_2$ has a mixing time of smaller order (it is w.h.p. a tree, and given that event, it is a uniform tree on its vertices and as such has $t_{\text{MIX}} \asymp |\mathcal{C}_2|^{3/2}$ (see e.g. [70]), that is $t_{\text{MIX}} \asymp \varepsilon^{-3} \log^{3/2}(\varepsilon^3 n)$ as $|\mathcal{C}_2| \asymp \varepsilon^{-2} \log(\varepsilon^3 n)$ w.h.p.), it turns out that w.h.p. there exists an even smaller component, whose mixing time is of the same order as on $\mathcal{C}_1$. This is captured by our second theorem, which also handles the subcritical regime.

Theorem 2 (controlling all components). *Let $G \sim \mathcal{G}(n, p)$ for $p = (1 \pm \varepsilon)/n$, where $\varepsilon \to 0$ and $\varepsilon^3 n \to \infty$. Let $\mathcal{C}^\star$ be the component of G that maximizes the mixing time of the lazy random walk on it, denoted by $t_{\text{MIX}}^\star$. Then with high probability, $t_{\text{MIX}}^\star$ has order $\varepsilon^{-3} \log^2(\varepsilon^3 n)$. This also holds when maximizing only over tree components.*

In the area of random walk on random graphs, the following two regimes have been analyzed extensively.

- The supercritical regime, where $t_{\text{MIX}} \asymp (\text{diam})^2$ with diam denoting the intrinsic diameter in the percolation cluster. Besides $\mathcal{G}(n, \frac{c}{n})$ for $c > 1$, this also holds in the torus $\mathbb{Z}_n^d$ by [7] and [67].

- The critical regime on a high dimensional torus, where $t_{\text{MIX}} \asymp (\text{diam})^3$. As mentioned above, for critical percolation on the complete graph, this was shown in [70]. For high dimensional tori, this is a consequence of [43].

To the best of our knowledge, our result is the first interpolation for the mixing time between these two different powers of the diameter.

Part I

Mixing time for the Ising model

Chapter 2

Mixing evolution of the mean-field Ising model

2.1 Introduction

The *Curie-Weiss* model is a special case of the Ising model (as in (1.1.1)) where the underlying geometry is the complete graph on n vertices. The study of this model (see, e.g., [31],[33],[34],[53]) is motivated by the fact that its behavior approximates that of the Ising model on high-dimensional tori. Throughout the chapter, we let $J \equiv 1$ and $H \equiv 0$ unless otherwise specified. It is convenient in this case to re-scale the parameter β, so that the stationary measure μ_n satisfies

$$\mu_n(\sigma) \propto \exp\left(\frac{\beta}{n}\sum_{x<y}\sigma(x)\sigma(y)\right) . \tag{2.1.1}$$

It is well known that for any fixed $\beta > 1$, the Glauber dynamics (X_t) mixes in exponential time (cf., e.g., [41]), whereas for any fixed $\beta < 1$ (high temperature) the mixing time has order $n \log n$ (see [1] and also [12]). Recently, Levin, Luczak and Peres [53] established that the mixing-time at the critical point $\beta = 1$ has order $n^{3/2}$, and that for fixed $0 < \beta < 1$ there is cutoff at time $\frac{1}{2(1-\beta)}n\log n$ with window n. It is therefore natural to ask how the phase transition between these states occurs around the critical $\beta_c = 1$: abrupt mixing at time $(\frac{1}{2(1-\beta)} + o(1))n\log n$ changes to a mixing-time of $\Theta(n^{3/2})$ steps, and finally to exponentially slow mixing.

In this chapter, we determine this phase transition, and characterize the mixing-time of the dynamics as a function of the parameter β, as it approaches its critical value $\beta_c = 1$ both from below and from above. The scaling window around the critical temperature β_c
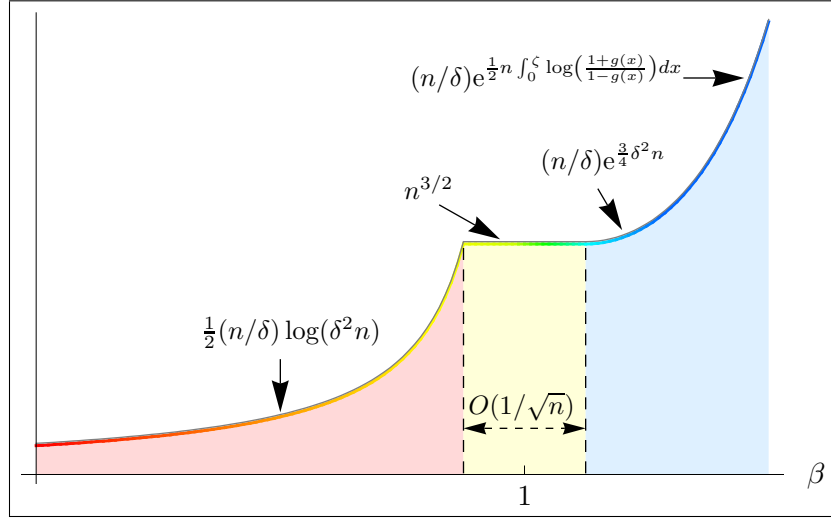
Figure 2.1: Illustration of the mixing time evolution as a function of the inverse-temperature β, with a scaling window of order $1/\sqrt{n}$ around the critical point. We write $\delta = |\beta - 1|$ and let ζ be the unique positive root of $g(x) := \frac{\tanh(\beta x) - x}{1 - x \tanh(\beta x)}$. Cutoff only occurs at high temperature.

has order $1/\sqrt{n}$, as formulated by the following theorems, and illustrated in Figure 2.1.

Theorem 3 (Subcritical regime). *Let $\delta = \delta(n) > 0$ be such that $\delta^2 n \to \infty$ with n. The Glauber dynamics for the mean-field Ising model with parameter $\beta = 1 - \delta$ exhibits cutoff at time $\frac{1}{2}(n/\delta) \log(\delta^2 n)$ with window size n/δ. In addition, the spectral gap of the dynamics in this regime is $(1 + o(1))\delta/n$, where the $o(1)$-term tends to 0 as $n \to \infty$.*

Theorem 4 (Critical window). *Let $\delta = \delta(n)$ satisfy $\delta = O(1/\sqrt{n})$. The mixing time of the Glauber dynamics for the mean-field Ising model with parameter $\beta = 1 \pm \delta$ has order $n^{3/2}$, and does not exhibit cutoff. In addition, the spectral gap of the dynamics in this regime has order $n^{-3/2}$.*

Theorem 5 (Supercritical regime). *Let $\delta = \delta(n) > 0$ be such that $\delta^2 n \to \infty$ with n. The mixing-time of the Glauber dynamics for the mean-field Ising model with parameter $\beta = 1 + \delta$ does not exhibit cutoff, and has order*

$$t_{\exp}(n) := \frac{n}{\delta} \exp\left(\frac{n}{2} \int_0^\zeta \log\left(\frac{1 + g(x)}{1 - g(x)}\right) dx\right) ,$$

where $g(x) := (\tanh(\beta x) - x) / (1 - x \tanh(\beta x))$, and ζ is the unique positive root of g. In particular, in the special case $\delta \to 0$, the order of the mixing time is $\frac{n}{\delta} \exp\left((\frac{3}{4} + o(1))\delta^2 n\right)$,

where the o(1)-term tends to 0 as n → ∞. In addition, the spectral gap of the dynamics in this regime has order $1/t_{\exp}(n)$.

As we further explain in Section 2.2, the key element in the proofs of the above theorems is understanding the behavior of the sum of all spins (known as the *magnetization chain*) at different temperatures. This function of the dynamics turns out to be an ergodic Markov chain as well, and namely a *birth-and-death* chain (a one-dimensional chain, where only moves between neighboring positions are permitted). In fact, the reason for the exponential mixing at low-temperature is essentially that this magnetization chain has two centers of mass, $\pm\zeta n$ (where ζ is as defined in Theorem 5), with an exponential commute time between them.

Recalling Theorem 3, the above confirms that there is a symmetric scaling window of order $1/\sqrt{n}$ around the critical temperature, beyond which there is cutoff both at high and at low temperatures, with the same order of mixing-time (yet with a different constant), cutoff window and spectral gap.

The rest of this chapter is organized as follows. Section 2.2 contains a brief outline of the proofs of the main theorems. Several preliminary facts on the Curie-Weiss model and on one-dimensional chains appear in Section 5.1. Sections 2.4, 2.5 and 2.6 address the high temperature regime (Theorem 3), critical temperature regime (Theorem 4) and low temperature regime (Theorem 5) respectively.

2.2 Outline of proof

In what follows, we present a sketch of the main ideas and arguments used in the proofs of the main theorems. We note that the analysis of the critical window relies on arguments similar to those used for the subcritical and supercritical regimes. Namely, to obtain the order of the mixing-time in Theorem 4 (critical window), we study the magnetization chain using the arguments that appear in the proof of Theorem 3 (high temperature regime). It is then straightforward to show that the mixing-time of the entire Glauber dynamics has the very same order. In turn, the spectral-gap in the critical window is obtained using arguments similar to those used in the proof of Theorem 5 (low temperature regime). In light of this, the following sketch will focus on the two non-critical temperature regimes.

2.2.1 High temperature regime

Upper bound for mixing

As mentioned above, a key element in the proof is the analysis of the *normalized magnetization* chain, (S_t), which is the average spin in the system. That is, for a given configuration σ,

we define $S(\sigma)$ to be $\frac{1}{n}\sum_i \sigma(i)$, and it is easy to verify that this function of the dynamics is an irreducible and aperiodic Markov chain. Clearly, a necessary condition for the mixing of the dynamics is the mixing of its magnetization, but interestingly, in our case the converse essentially holds as well. For instance, as we later explain, in the special case where the starting state is the all-plus configuration, by symmetry these two chains have precisely the same total variation distance from equilibrium at any given time.

In order to determine the behavior of the chain (S_t), we first keep track of its expected value along the Glauber dynamics. To simplify the sketch of the argument, suppose that our starting configuration is somewhere near the all-plus configuration. In this case, one can show that $\mathbf{E}S_t$ is monotone decreasing in t, and drops to order $\sqrt{1/\delta n}$ precisely at the cutoff point. Moreover, if we allow the dynamics to perform another $\Theta(n/\delta)$ steps (our cutoff window), then the magnetization will hit 0 (or $\frac{1}{n}$, depending on the parity of n) with probability arbitrarily close to 1. At that point, we essentially achieve the mixing of the magnetization chain.

It remains to extend the mixing of the magnetization chain to the mixing of the entire Glauber dynamics. Roughly, keeping in mind the above comment on the symmetric case of the all-plus starting configuration, one can apply a similar argument to an arbitrary starting configuration σ, by separately treating the set of spins which were initially positive and those which were initially negative. Indeed, it was shown in [53] that the following holds for $\beta < 1$ fixed (strictly subcritical regime). After a "burn-in" period of order n steps, the magnetization typically becomes not too biased. Next, if one runs two instances of the dynamics, from two such starting configurations (where the magnetization is not too biased), then by the time it takes their magnetization chains to coalesce, the entire configurations become relatively similar. This was established by a so-called *Two Coordinate Chain* analysis, where the two coordinates correspond to the current sum of spins along the set of sites which were initially either positive or negative respectively.

By extending the above Two Coordinate Chain Theorem to the case of $\beta = 1 - \delta$ where $\delta = \delta(n)$ satisfies $\delta^2 n \to \infty$, and combining it with second moment arguments and some additional ideas, we were able to show that the above behavior holds throughout this mildly subcritical regime. The burn-in time required for the typical magnetization to become "balanced" now has order n/δ, and so does the time it takes the full dynamics of two chains to coalesce once their magnetization chains have coalesced. Thus, these two periods are conveniently absorbed in our cutoff window, making the cutoff of the magnetization chain the dominant factor in the mixing of the entire Glauber dynamics.

Lower bound for mixing

While the above mentioned Two Coordinate Chain analysis was required in order to show that the entire Glauber dynamics mixes fairly quickly once its magnetization chain reaches equilibrium, the converse is immediate. Thus, we will deduce the lower bound on the mixing time of the dynamics solely from its magnetization chain.

The upper bound in this regime relied on an analysis of the first and second moments of the magnetization chain, however this approach is too coarse to provide a precise lower bound for the cutoff. We therefore resort to establishing an upper bound on the *third* moment of the magnetization chain, using which we are able to fine-tune our analysis of how its first moment changes along time. Examining the state of the system order n/δ steps before the alleged cutoff point, using concentration inequalities, we show that the magnetization chain is typically substantially far from 0. This implies a lower bound on the total variation distance of the magnetization chain to stationarity, as required.

Spectral gap analysis

In the previous arguments, we stated that the magnetization chain essentially dominates the mixing-time of the entire dynamics. An even stronger statement holds for the spectral gap: the Glauber dynamics and its magnetization chain have precisely the same spectral gap, and it is in both cases attained by the second largest eigenvalue. We therefore turn to establish the spectral gap of (S_t).

The lower bound follows directly from the contraction properties of the chain in this regime. To obtain a matching upper bound, we use the Dirichlet representation for the spectral gap, combined with an appropriate bound on the *fourth* moment of the magnetization chain.

2.2.2   Low temperature regime

Exponential mixing

As mentioned above, the exponential mixing in this regime follows directly from the behavior of the magnetization chain, which has a bottleneck between $\pm\zeta$. To show this, we analyze the effective resistance between these two centers of mass, and obtain the precise order of the commute time between them. Additional arguments show that the mixing time of the entire Glauber dynamics in this regime has the same order.

Spectral gap analysis

In the above mentioned proof of the exponential mixing, we establish that the commute time of the magnetization chain between 0 and ζ has the same order as the hitting time from 1 to 0. We can therefore apply a recent result of [25] for general birth-and-death chains, which implies that in this case the inverse of the spectral-gap (known as the relaxation-time) and the mixing-time must have the same order.

2.3 Preliminaries

2.3.1 The magnetization chain

The normalized magnetization of a configuration $\sigma \in \Omega$, denoted by $S(\sigma)$, is defined as

$$S(\sigma) := \frac{1}{n} \sum_{i=1}^{n} \sigma(i) \; .$$

Suppose that the current state of the Glauber dynamics is σ, and that site i has been selected to have its spin updated. By definition, the probability of updating this site to a positive spin is given by $p^+ \left(S(\sigma) - \sigma(i)/n \right)$, where

$$p^+(s) := \frac{e^{\beta s}}{e^{\beta s} + e^{-\beta s}} = \frac{1 + \tanh(\beta s)}{2} \; . \tag{2.3.1}$$

Similarly, the probability of updating the spin of site i to a negative one is given by $p^- \left(S(\sigma) - \sigma(i)/n \right)$, where

$$p^-(s) := \frac{e^{-\beta s}}{e^{\beta s} + e^{-\beta s}} = \frac{1 - \tanh(\beta s)}{2} \; . \tag{2.3.2}$$

It follows that the (normalized) magnetization of the Glauber dynamics at each step is a Markov chain, (S_t), with the following transition kernel:

$$P_M(s, s') = \begin{cases} \frac{1+s}{2} p^-(s - n^{-1}) & \text{if } s' = s - \frac{2}{n}, \\ \frac{1-s}{2} p^+(s + n^{-1}) & \text{if } s' = s + \frac{2}{n}, \\ 1 - \frac{1+s}{2} p^-(s - n^{-1}) - \frac{1-s}{2} p^+(s + n^{-1}) & \text{if } s' = s \; . \end{cases} \tag{2.3.3}$$

An immediate important property that the above reveals is the symmetry of S_t: the distribution of $(S_{t+1} \mid S_t = s)$ is precisely that of $(-S_{t+1} \mid S_t = -s)$.

As evident from the above transition rules, the behavior of the Hyperbolic tangent will be useful in many arguments. This is illustrated in the following simple calculation, showing that the minimum over the holding probabilities of the magnetization chain is nearly $\frac{1}{2}$. Indeed, since the derivative of $\tanh(x)$ is bounded away from 0 and 1 for all $x \in [0, \beta]$ and any $\beta = O(1)$, the Mean Value Theorem gives

$$
\begin{array}{rcl}
P_M(s, s + \frac{2}{n}) & = & \frac{1-s}{4}\left(1 + \tanh(\beta s)\right) + O(n^{-1}) , \\
P_M(s, s - \frac{2}{n}) & = & \frac{1+s}{4}\left(1 - \tanh(\beta s)\right) + O(n^{-1}) , \\
P_M(s, s) & = & \frac{1}{2}\left(1 + s\tanh(\beta s)\right) - O(n^{-1}) .
\end{array}
\tag{2.3.4}
$$

Therefore, the holding probability in state s is at least $\frac{1}{2} - O\left(\frac{1}{n}\right)$. In fact, since $\tanh(x)$ is monotone increasing, $P_M(s, s) \leq \frac{1}{2} + \frac{1}{2}s\tanh(\beta s)$ for all s, hence these probability are also bounded from above by $\frac{1}{2}(1 + \tanh(\beta)) < 1$.

Using the above fact, the next lemma will provide an upper bound for the coalescence time of two magnetization chains, S_t and $\tilde{S}_t$, in terms of the hitting time τ_0, defined as $\tau_0 := \min\{t : |S_t| \leq n^{-1}\}$.

Lemma 2.3.1. *Let (S_t) and $(\tilde{S}_t)$ denote two magnetization chains, started from two arbitrary states. Then for any $\varepsilon > 0$ there exists some $c_\varepsilon > 0$, such that the following holds: if $T > 0$ satisfies $\mathbf{P}_1(\tau_0 \geq T) < \varepsilon$ then S_t and $\tilde{S}_t$ can be coupled in a way such that they coalesce within at most $c_\varepsilon T$ steps with probability at least $1 - \varepsilon$.*

Proof. Assume without loss of generality that $|\tilde{S}_0| < |S_0|$, and by symmetry, that $\sigma = |S_0| \geq 0$. Define

$$
\tau := \min\left\{t : |S_t| \leq |\tilde{S}_t| + \frac{2}{n}\right\} .
$$

Recalling the definition of τ_0, clearly we must have $\tau < \tau_0$. Next, since the holding probability of S_t at any state s is bounded away from 0 and 1 for large n (by the discussion preceding the lemma), there clearly exists a constant $0 < b < 1$ such that

$$
\mathbf{P}\left(S_{t+1} = \tilde{S}_{t+1} \,\middle|\, |S_t - \tilde{S}_t| \leq \frac{2}{n}\right) > b > 0
$$

(for instance, one may choose $b = \frac{1}{10}(1 - \tanh(\beta))$ for a sufficiently large n). It therefore follows that $|S_{\tau+1}| = |\tilde{S}_{\tau+1}|$ with probability at least b.

Condition on this event. We claim that in this case, the coalescence of (S_t) and $(\tilde{S}_t)$ (rather than just their absolute values) occurs at some $t \leq \tau_0 + 1$ with probability at least b. The case $S_{\tau+1} = \tilde{S}_{\tau+1}$ is immediate, and it remains to deal with the case $S_{\tau+1} = -\tilde{S}_{\tau+1}$. Let us couple (S_t) and $(\tilde{S}_t)$ so that the property $S_t = -\tilde{S}_t$ is maintained henceforth. Thus, at time $t = \tau_0$ we obtain $|S_t - \tilde{S}_t| = 2|S_t| \leq \frac{2}{n}$, and with probability b this yields $S_{t+1} = \tilde{S}_{t+1}$.

Clearly, our assumption on T and the fact that $0 \leq \sigma \leq 1$ together give

$$\mathbf{P}_\sigma(\tau_0 \geq T) \leq \mathbf{P}_1(\tau_0 \geq T) < \varepsilon .$$

Thus, with probability at least $(1 - \varepsilon)b^2$, the coalescence time of (S_t) and $(\tilde{S}_t)$ is at most T. Repeating this experiment a sufficiently large number of times then completes the proof. ∎

In order to establish cutoff for the magnetization chain (S_t), we will need to carefully track its moments along the Glauber dynamics. By definition (see (2.3.3)), the behavior of these moments is governed by the Hyperbolic tangent function, as demonstrated by the following useful form for the conditional expectation of S_{t+1} given S_t (see also [53, (2.13)]).

$$\begin{aligned}
\mathbf{E}\left[S_{t+1} \mid S_t = s\right] &= \left(s + \tfrac{2}{n}\right)P_M\left(s, s + \tfrac{2}{n}\right) + sP_M(s, s) + \left(s - \tfrac{2}{n}\right)P_M\left(s, s - \tfrac{2}{n}\right) \\
&= (1 - n^{-1})s + \varphi(s) - \psi(s) ,
\end{aligned} \tag{2.3.5}$$

where

$$\varphi(s) = \varphi(s, \beta, n) := \frac{1}{2n}\left[\tanh\left(\beta(s + n^{-1})\right) + \tanh\left(\beta(s - n^{-1})\right)\right] ,$$

$$\psi(s) = \psi(s, \beta, n) := \frac{s}{2n}\left[\tanh\left(\beta(s + n^{-1})\right) - \tanh\left(\beta(s - n^{-1})\right)\right] .$$

2.3.2 From magnetization equilibrium to full mixing

The motivation for studying the magnetization chain is that its mixing essentially dominates the full mixing of the Glauber dynamics. This is demonstrated by the next straightforward lemma (see also [53, Lemma 3.4]), which shows that in the special case where the starting point is the all-plus configuration, the mixing of the magnetization is *precisely* equivalent to that of the entire dynamics.

Lemma 2.3.2. *Let (X_t) be an instance of the Glauber dynamics for the mean field Ising model starting from the all-plus configuration, namely, $\sigma_0 = \mathbf{1}$, and let $S_t = S(X_t)$ be its magnetization chain. Then*

$$\|\mathbf{P}_{\mathbf{1}}(X_t \in \cdot) - \mu_n\|_{\mathrm{TV}} = \|\mathbf{P}_{\mathbf{1}}(S_t \in \cdot) - \pi_n\|_{\mathrm{TV}} , \tag{2.3.6}$$

where π_n is the stationary distribution of the magnetization chain.

Proof. For any $s \in \{-1, -1 + \tfrac{2}{n}, \ldots, 1 - \tfrac{2}{n}, 1\}$, let $\Omega_s := \{\sigma \in \Omega : S(\sigma) = s\}$. Since by symmetry, both $\mu_n(\cdot \mid \Omega_s)$ and $\mathbf{P}_{\mathbf{1}}(X_t \in \cdot \mid S_t = s)$ are uniformly distributed over Ω_s, the

following holds:

$$\|\mathbf{P_1}(X_t \in \cdot) - \mu_n\|_{\text{TV}} = \frac{1}{2} \sum_s \sum_{\sigma \in \Omega_s} |\mathbf{P_1}(X_t = \sigma) - \mu_n(\sigma)|$$

$$= \frac{1}{2} \sum_s \sum_{\sigma \in \Omega_s} \left| \frac{\mathbf{P_1}(S_t = s)}{|\Omega_s|} - \frac{\mu_n(\Omega_s)}{|\Omega_s|} \right|$$

$$= \|\mathbf{P_1}(S_t \in \cdot) - \pi_n\|_{\text{TV}} . \qquad \blacksquare$$

In the general case where the Glauber dynamics starts from an arbitrary configuration σ_0, though the above equivalence (2.3.6) no longer holds, the magnetization still dominates the full mixing of the dynamics in the following sense. The full coalescence of two instances of the dynamics occurs within order $n \log n$ steps once the magnetization chains have coalesced.

Lemma 2.3.3 ([53, Lemma 2.9]). *Let $\sigma, \tilde{\sigma} \in \Omega$ be such that $S(\sigma) = S(\tilde{\sigma})$. For a coupling $(X_t, \tilde{X}_t)$, define the coupling time $\tau_{X,\tilde{X}} := \min\{t \geq 0 : X_t = \tilde{X}_t\}$. Then for a sufficiently large $c_0 > 0$ there exists a coupling $(X_t, \tilde{X}_t)$ of the Glauber dynamics with initial states $X_0 = \sigma$ and $\tilde{X}_0 = \tilde{\sigma}$ such that*

$$\limsup_{n \to \infty} \mathbf{P}_{\sigma,\tilde{\sigma}} \left(\tau_{X,\tilde{X}} > c_0 n \log n \right) = 0 .$$

Though Lemma 2.3.3 holds for any temperature, it will only prove useful in the critical and low temperature regimes. At high temperature, using more delicate arguments, we will establish full mixing within order of $\frac{n}{\delta}$ steps once the magnetization chains have coalesced. That is, the extra steps required to achieve full mixing, once the magnetization chain cutoff had occurred, are absorbed in the cutoff window. Thus, in this regime, the entire dynamics has cutoff precisely when its magnetization chain does (with the same window).

2.3.3   Contraction and one-dimensional Markov chains

We say that a Markov chain, assuming values in $\mathbb{R}$, is *contracting*, if the expected distance between two chains after a single step decreases by some factor bounded away from 0. As we later show, the magnetization chain is contracting at high temperatures, a fact which will have several useful consequences. One example of this is the following straightforward lemma of [53], which provides a bound on the variance of the chain. Here and throughout the chapter, the notation $\mathbf{P}_z$, $\mathbf{E}_z$ and Var_z will denote the probability, expectation and variance respectively given that the starting state is z.

Lemma 2.3.4 ([53, Lemma 2.6]). *Let (Z_t) be a Markov chain taking values in $\mathbb{R}$ and with transition matrix P. Suppose that there is some $0 < \rho < 1$ such that for all pairs of starting*

states $(z, \tilde{z})$,

$$| \mathbf{E}_z[Z_t] - \mathbf{E}_{\tilde{z}}[Z_t] | \leq \rho^t |z - \tilde{z}|. \tag{2.3.7}$$

Then $v_t := \sup_{z_0} \mathrm{Var}_{z_0}(Z_t)$ *satisfies* $v_t \leq v_1 \min \{t, 1/(1 - \rho^2)\}$.

Remark. By following the original proof of the above lemma, one can readily extend it to the case $\rho \geq 1$ and get the following bound:

$$v_t \leq v_1 \cdot \rho^{2t} \min \{t, 1/(\rho^2 - 1)\} . \tag{2.3.8}$$

This bound will prove to be effective for reasonably small values of t in the critical window, where although the magnetization chain is not contracting, ρ is only slightly larger than 1.

Another useful property of the magnetization chain in the high temperature regime is its drift towards 0. As we later show, in this regime, for any $s > 0$ we have $\mathbf{E}[S_{t+1}|S_t = s] < s$, and with probability bounded below by a constant we have $S_{t+1} < S_t$. We thus refer to the following lemma of [54]:

Lemma 2.3.5 ([54, Chapter 18]). *Let* $(W_t)_{t \geq 0}$ *be a non-negative supermartingale and* τ *be a stopping time such*

(i) $W_0 = k$,

(ii) $W_{t+1} - W_t \leq B$,

(iii) $\mathrm{Var}(W_{t+1} \mid \mathcal{F}_t) > \sigma^2 > 0$ *on the event* $\tau > t$.

If $u > 4B^2/(3\sigma^2)$, *then* $\mathbf{P}_k(\tau > u) \leq \frac{4k}{\sigma\sqrt{u}}$.

This lemma, together with the above mentioned properties of (S_t), yields the following immediate corollary:

Corollary 2.3.6 ([53, Lemma 2.5]). *Let* $\beta \leq 1$, *and suppose that* n *is even. There exists a constant* c *such that, for all* s *and for all* $u, t \geq 0$,

$$\mathbf{P}(|S_u| > 0, \ldots, |S_{u+t}| > 0 \mid S_u = s) \leq \frac{cn|s|}{\sqrt{t}} . \tag{2.3.9}$$

Finally, our analysis of the spectral gap of the magnetization chain will require several results concerning birth-and-death chains from [25]. In what follows and throughout the chapter, the relaxation-time of a chain, t_{REL}, is defined to be $\mathtt{gap}^{-1}$, where $\mathtt{gap}$ denotes its spectral-gap. We say that a chain is *b-lazy* if all its holding probabilities are at least b, or simply *lazy* for the useful case of $b = \frac{1}{2}$. Finally, given an ergodic birth-and-death chain on

$\mathcal{X} = \{0, 1, \ldots, n\}$ with stationary distribution π, the *quantile state* $Q(\alpha)$, for $0 < \alpha < 1$, is defined to be the smallest $i \in \mathcal{X}$ such that $\pi(\{0, \ldots, i\}) \geq \alpha$.

Lemma 2.3.7 ([25, Lemma 2.9]). *Let $X(t)$ be a lazy irreducible birth-and-death chain on $\{0, 1, \ldots, n\}$, and suppose that for some $0 < \varepsilon < \frac{1}{16}$ we have $t_{\mathrm{REL}} < \varepsilon^4 \cdot \mathbf{E}_0 \tau_{Q(1-\varepsilon)}$. Then for any fixed $\varepsilon \leq \alpha < \beta \leq 1 - \varepsilon$:*

$$\mathbf{E}_{Q(\alpha)} \tau_{Q(\beta)} \leq \frac{3}{2\varepsilon} \sqrt{t_{\mathrm{REL}} \cdot \mathbf{E}_0 \tau_{Q(\frac{1}{2})}} \ . \tag{2.3.10}$$

Lemma 2.3.8 ([25, Lemma 2.3]). *For any fixed $0 < \varepsilon < 1$ and lazy irreducible birth-and-death chain X, the following holds for any t:*

$$\|P^t(0, \cdot) - \pi\|_{\mathrm{TV}} \leq \mathbf{P}_0(\tau_{Q(1-\varepsilon)} > t) + \varepsilon \ , \tag{2.3.11}$$

and for all $k \in \Omega$,

$$\|P^t(k, \cdot) - \pi\|_{\mathrm{TV}} \leq \mathbf{P}_k(\max\{\tau_{Q(\varepsilon)}, \tau_{Q(1-\varepsilon)}\} > t) + 2\varepsilon \ . \tag{2.3.12}$$

Remark. As argued in [25] (see Theorem 3.1 and its proof), the above two lemmas also hold for the case where the birth-and-death chain is not lazy but rather b-lazy for some constant $b > 0$. The formulation for this more general case incurs a cost of a slightly different constant in (2.3.10), and replacing t with t/C (for some constant C) in (2.3.11) and (2.3.12). As we already established (recall (2.3.4)), the magnetization chain is indeed b-lazy for any constant $b < \frac{1}{2}$ and a sufficiently large n.

2.3.4 Monotone coupling

A useful tool throughout our arguments is the *monotone coupling* of two instances of the Glauber dynamics (X_t) and $(\tilde{X}_t)$, which maintains a coordinate-wise inequality between the corresponding configurations. That is, given two configurations $\sigma \geq \tilde{\sigma}$ (i.e., $\sigma(i) \geq \tilde{\sigma}(i)$ for all i), it is possible to generate the next two states σ' and $\tilde{\sigma}'$ by updating the same site in both, in a manner that ensures that $\sigma' \geq \tilde{\sigma}'$. More precisely, we draw a random variable I uniformly over $\{1, 2, \ldots, n\}$ and independently draw another random variable U uniformly over $[0, 1]$. To generate σ' from σ, we update site I to $+1$ if $U \leq p^+ \left(S(\sigma) - \frac{\sigma(I)}{n} \right)$, otherwise $\sigma'(I) = -1$. We perform an analogous process in order to generate $\tilde{\sigma}'$ from $\tilde{\sigma}$, using the same I and U as before. The monotonicity of the function p^+ guarantees that $\sigma' \geq \tilde{\sigma}'$, and by repeating this process, we obtain a coupling of the two instances of the Glauber dynamics that always maintains monotonicity.

Clearly, the above coupling induces a monotone coupling for the two corresponding magnetization chains. We say that a birth-and-death chain with a transition kernel P and a state-space $\mathcal{X} = \{0, 1, \ldots, n\}$ is *monotone* if $P(i, i+1) + P(i+1, i) \le 1$ for every $i < n$. It is easy to verify that this condition is equivalent to the existence of a monotone coupling, and that for such a chain, if $f : \mathcal{X} \to \mathbb{R}$ is a monotone increasing (decreasing) function then so is Pf (see, e.g., [25, Lemma 4.1]).

2.3.5 The spectral gap of the dynamics and its magnetization chain

To analyze the spectral gap of the Glauber dynamics, we establish the following lemma which reduces this problem to determining the spectral-gap of the one-dimensional magnetization chain. Its proof relies on increasing eigenfunctions, following the ideas of [71].

Proposition 2.3.9. *The Glauber dynamics for the mean-field Ising model and its one-dimensional magnetization chain have the same spectral gap. Furthermore, both gaps are attained by the largest nontrivial eigenvalue.*

Proof. We will first show that the one-dimensional magnetization chain has an increasing eigenfunction, corresponding to the second eigenvalue.

Recalling that S_t assumes values in $\mathcal{X} := \{-1, -1 + \frac{2}{n}, \ldots, 1 - \frac{2}{n}, 1\}$, let M denote its transition matrix, and let π denote its stationary distribution. Let $1 = \theta_0 \ge \theta_1 \ge \ldots \ge \theta_n$ be the $n+1$ eigenvalues of M, corresponding to the eigenfunctions $f_0 \equiv 1, f_1, \ldots, f_n$. Define $\theta = \max\{\theta_1, |\theta_n|\}$, and notice that, as S_t is aperiodic and irreducible, $0 < \theta < 1$. Furthermore, by the existence of the monotone coupling for S_t and the discussion in the previous subsection, whenever a function $f : \mathcal{X} \to \mathbb{R}$ is increasing so is Mf.

Define $f : I \to \mathbb{R}$ by $f := f_1 + f_n + K\mathbf{1}$, where $\mathbf{1}$ is the identity function and $K > 0$ is sufficiently large to ensure that f is monotone increasing (e.g., $K = \frac{n}{2}\|f_1 + f_n\|_{L^\infty}$ easily suffices). Notice that, by symmetry of S_t, $\pi(x) = \pi(-x)$ for all $x \in \mathcal{X}$, and in particular $\sum_{x \in \mathcal{X}} x\pi(x) = 0$, that is to say, $\langle \mathbf{1}, f_0 \rangle_{L^2(\pi)} = 0$. Recalling that for all $i \ne j$ we have $\langle f_i, f_j \rangle_{L^2(\pi)} = 0$, it follows that for some $q_1, \ldots, q_n \in \mathbb{R}$ we have $f = \sum_{i=1}^n q_i f_i$ with $q_1 \ne 0$ and $q_n \ne 0$, and thus

$$\left(\theta^{-1}M\right)^m f = \sum_{i=1}^n q_i (\theta_i/\theta)^m f_i \ .$$

Next, define

$$g = \begin{cases} q_1 f_1 & \text{if } \theta = \theta_1 \\ 0 & \text{otherwise} \end{cases} \ , \quad \text{and} \quad h = \begin{cases} q_n f_n & \text{if } \theta = -\theta_n \\ 0 & \text{otherwise} \end{cases} \ ,$$

and notice that

$$\lim_{m\to\infty} \left(\theta^{-1}M\right)^{2m} f = g + h , \quad \text{and} \quad \lim_{m\to\infty} \left(\theta^{-1}M\right)^{2m+1} f = g - h .$$

As stated above, $M^m f$ is increasing for all m, and thus so are the two limits $g+h$ and $g-h$ above, as well as their sum. We deduce that g is an increasing function, and next claim that $g \not\equiv 0$. Indeed, if $g \equiv 0$ then both h and $-h$ are increasing functions, hence necessarily $h \equiv 0$ as well; this would imply that $q_1 = q_n = 0$, thus contradicting our construction of f.

We deduce that g is an increasing eigenfunction corresponding to $\theta_1 = \theta$, and next wish to show that it is strictly increasing. Recall that for all $x \in \mathcal{X}$,

$$(Mg)(x) = M\left(x, x - \frac{2}{n}\right)g\left(x - \frac{2}{n}\right) + M(x,x)g(x) + M\left(x, x + \frac{2}{n}\right)g\left(x + \frac{2}{n}\right) .$$

Therefore, if for some $x \in \mathcal{X}$ we had $g(x - \frac{2}{n}) = g(x) \geq 0$, the fact that g is increasing would imply that

$$\theta_1 g(x) = (Mg)(x) \geq g(x) \geq 0 ,$$

and analogously, if $g(x) = g(x + \frac{2}{n}) \leq 0$ we could write

$$\theta_1 g(x) = (Mg)(x) \leq g(x) \leq 0 .$$

In either case, since $0 < \theta_1 < 1$ (recall that $\theta_1 = \theta$) this would in turn lead to $g(x) = 0$. By inductively substituting this fact in the above equation for $(Mg)(x)$, we would immediately get $g \equiv 0$, a contradiction.

Let $1 = \lambda_0 \geq \lambda_1 \geq \ldots \geq \lambda_{|\Omega|-1}$ denote the eigenvalues of the Glauber dynamics, and let $\lambda := \max\{\lambda_1, |\lambda_{2^n-1}|\}$. We translate g into a function $G : \Omega \to \mathbb{R}$ in the obvious manner:

$$G(\sigma) := g(S(\sigma)) = g\left(\frac{1}{n}\sum_{i=1}^{n} \sigma(i)\right) .$$

One can verify that G is indeed an eigenfunction of the Glauber dynamics corresponding to the eigenvalue θ_1, and clearly G is strictly increasing with respect to the coordinate-wise partial order on Ω. At this point, we refer to the following lemma of [71]:

Lemma 2.3.10 ([71, Lemma 4]). *Let P be the transition matrix of the Glauber dynamics, and let λ_1 be its second largest eigenvalue. If P has a strictly increasing eigenfunction f, then f corresponds to λ_1.*

The above lemma immediately implies that G corresponds to the second eigenvalue of Glauber dynamics, which we denote by λ_1, and thus $\lambda_1 = \theta_1$.

It remains to show that $\lambda = \lambda_1$. To see this, first recall that all the holding probabilities of S_t are bounded away from 0, and the same applies to the entire Glauber dynamics by definition (the magnetization remains the same if and only if the configuration remains the same). Therefore, both θ_n and λ_{2^n-1} are bounded away from -1, and it remains to show that $\mathtt{gap} = o(1)$ for the Glauber dynamics (and hence also for its magnetization chain).

To see this, suppose P is the transition kernel of the Glauber dynamics, and recall the Dirichlet representation for the second eigenvalue of a reversible chain (see [54, Lemma 13.7], and also [3, Chapter 3]):

$$1 - \lambda_1 = \min\left\{ \frac{\mathcal{E}(f)}{\langle f, f\rangle_{\mu_n}} \; : \; f \not\equiv 0 \, , \; \mathbf{E}_{\mu_n}(f) = 0 \right\} , \tag{2.3.13}$$

where $\mathbf{E}_{\mu_n}(f)$ denotes $\langle \mathbf{1}, f\rangle_{\mu_n}$, and

$$\mathcal{E}(f) = \langle (I - P)f, f\rangle_{\mu_n} = \frac{1}{2} \sum_{\sigma, \sigma' \in \Omega} [f(\sigma) - f(\sigma')]^2 \, \mu_n(\sigma) P(\sigma, \sigma') .$$

By considering the sum of spins, $h(\sigma) = \sum_{i=1}^n \sigma(i)$, we get $\mathcal{E}(h) \leq 2$, and since the spins are positively correlated, $\mathrm{Var}_{\mu_n} \sum_i \sigma(i) \geq n$. It follows that

$$1 - \lambda_1 \leq 2/n ,$$

and thus $\mathtt{gap} = 1 - \lambda_1 = 1 - \theta_1$ for both the Glauber dynamics and its magnetization chain, as required. $\blacksquare$

2.4   High temperature regime

In this section we prove Theorem 3. Subsection 2.4.1 establishes the cutoff of the magnetization chain, which immediately provides a lower bound on the mixing time of the entire dynamics. The matching upper bound, which completes the proof of cutoff for the Glauber dynamics, is given in Subsection 2.4.2. The spectral gap analysis appears in Subsection 2.4.3. Unless stated otherwise, assume throughout this section that $\beta = 1 - \delta$ where $\delta^2 n \to \infty$.

2.4.1   Cutoff for the magnetization chain

Clearly, the mixing of the Glauber dynamics ensures the mixing of its magnetization. Interestingly, the converse is also essentially true, as the mixing of the magnetization turns out to be the most significant part in the mixing of the full Glauber dynamics. We thus wish to

prove the following cutoff result:

Theorem 2.4.1. *Let $\beta = 1 - \delta$, where $\delta > 0$ satisfies $\delta^2 n \to \infty$. Then the corresponding magnetization chain (S_t) exhibits cutoff at time $\frac{1}{2} \cdot \frac{n}{\delta} \log(\delta^2 n)$ with a window of order n/δ.*

Notice that Lemma 2.3.2 then gives the following corollary for the special case where the initial state of the dynamics is the all-plus configuration:

Corollary 2.4.2. *Let $\delta = \delta(n) > 0$ be such that $\delta^2 n \to \infty$ with n, and let (X_t) denote the Glauber dynamics for the mean-field Ising model with parameter $\beta = 1 - \delta$, started from the all-plus configuration. Then (X_t) exhibits cutoff at time $\frac{1}{2}(n/\delta) \log(\delta^2 n)$ with window size n/δ.*

Upper bound

Our goal in this subsection is to show the following:

$$\lim_{\gamma \to \infty} \limsup_{n \to \infty} d_n \left(\frac{1}{2} \cdot \frac{n}{\delta} \log(\delta^2 n) + \gamma \frac{n}{\delta} \right) = 0 \;, \tag{2.4.1}$$

where $d_n(\cdot)$ is with respect to the magnetization chain (S_t) and its stationary distribution. This will be obtained using an upper bound on the coalescence time of two instances of the magnetization chain. Given the properties of its stationary distribution (see Figure **??**), we will mainly be interested in the time it takes this chain to hit near 0. The following theorem provides an upper bound for that hitting time.

Theorem 2.4.3. *For $0 < \beta < 1 + O(n^{-1/2})$, consider the magnetization chain started from some arbitrary state s_0, and let $\tau_0 = \min\{t : |S_t| \leq n^{-1}\}$. Write $\beta = 1 - \delta$, and for $\gamma > 0$ define*

$$t_n(\gamma) = \begin{cases} \dfrac{n}{2\delta} \log(\delta^2 n) + (\gamma + 3)\dfrac{n}{\delta} & \delta^2 n \to \infty \;, \\[2mm] \left(200 + 6\gamma \left(1 + 6\sqrt{\delta^2 n} \right) \right) n^{3/2} & \delta^2 n = O(1) \;. \end{cases} \tag{2.4.2}$$

Then there exists some $c > 0$ such that $\mathbf{P}_{s_0}(\tau_0 > t_n(\gamma)) \leq c/\sqrt{\gamma}$.

Proof. For any $t \geq 1$, define:

$$s_t := \mathbf{E}_{s_0} \left[|S_t| \mathbf{1}_{\{\tau_0 > t\}} \right] \;.$$

Suppose $s > 0$. Recalling (2.3.5) and bearing in mind the concavity of the Hyperbolic tangent and the fact that $\psi(s) \geq 0$, we obtain that

$$\mathbf{E}(S_{t+1} \,|\, S_t = s) \leq s + \frac{1}{n} \left(\tanh(\beta s) - s \right) \;.$$

Using symmetry for the case $s < 0$, we can then deduce that

$$\mathbf{E}\left[|S_{t+1}| \,\big|\, S_t\right] \le |S_t| + \frac{1}{n}\left(\tanh(\beta|S_t|) - |S_t|\right) \text{ for any } t < \tau_0 \ . \tag{2.4.3}$$

Hence, combining the concavity of the Hyperbolic tangent together with Jensen's inequality yields

$$s_{t+1} \le \left(1 - \frac{1}{n}\right) s_t + \frac{1}{n}\tanh(\beta s_t) \ . \tag{2.4.4}$$

Since the Taylor expansion of $\tanh(x)$ is

$$\tanh(x) = x - \frac{x^3}{3} + \frac{2x^5}{15} - \frac{17x^7}{315} + O(x^9) \ , \tag{2.4.5}$$

we have $\tanh(x) \le x - \frac{x^3}{5}$ for $0 \le x \le 1$, giving

$$s_{t+1} \le \left(1 - \frac{1}{n}\right) s_t + \frac{1}{n}\tanh(\beta s_t) \le \left(1 - \frac{1}{n}\right) s_t + \frac{1}{n}\beta s_t - \frac{(s_t)^3}{5n}$$
$$= s_t - \frac{\delta}{n}s_t - \frac{(s_t)^3}{5n} \ . \tag{2.4.6}$$

For some $1 < a \le 2$ to be defined later, set

$$b_i = a^{-i}/4 \ , \text{ and } u_i = \min\{t : s_t \le b_i\} \ .$$

Notice that s_t is decreasing in t by (2.4.6), thus for every $t \in [u_i, u_{i+1}]$ we have

$$b_i/a = b_{i+1} \le s_t \le b_i \ .$$

It follows that

$$s_{t+1} \le s_t - \frac{\delta}{n} \cdot \frac{b_i}{a} - \frac{b_i^3}{5a^3 n} \ ,$$

and

$$u_{i+1} - u_i \le \left(\frac{a-1}{a}b_i\right) \Big/ \left(\frac{\delta}{n}\frac{b_i}{a} + \frac{b_i^3}{5a^3 n}\right) \le \frac{5(a-1)a^2 n}{5\delta a^2 + b_i^2} \ . \tag{2.4.7}$$

For the case $\delta^2 n \to \infty$, define:

$$i_0 = \min\{i : b_i \le 1/\sqrt{\delta n}\} \ .$$

The following holds:

$$\sum_{i=1}^{i_0}(u_{i+1}-u_i) \le \sum_{i=1}^{i_0}\frac{5(a-1)a^2n}{5\delta a^2 + b_i^2} \le \sum_{\substack{i\le i_0 \\ b_i^2 > \delta}}\frac{5(a-1)a^2n}{b_i^2} + \sum_{\substack{i\le i_0 \\ b_i^2 \le \delta}}\frac{5(a-1)a^2n}{5\delta a^2}$$

$$\le \frac{5na^2}{\delta(a+1)} + \frac{a-1}{2\log a}\cdot\frac{n}{\delta}\log(\delta^2 n)\ ,$$

where in the last inequality we used the fact that the series $\{b_i^{-2}\}$ is a geometric series with a ratio a^2, and that, as $b_i^2 \ge 1/(\delta n)$ for all $i \le i_0$, the number of summands such that $b_i^2 \le \delta$ is at most $\log_a(\sqrt{\delta^2 n})$. Therefore, choosing $a = 1 + n^{-1}$, we deduce that:

$$\sum_{i=1}^{i_0}(u_{i+1}-u_i) \le \left(\frac{5}{2} + O(n^{-1})\right)\frac{n}{\delta} + \left(\frac{1}{2} + O(n^{-1})\right)\frac{n}{\delta}\log(\delta^2 n)$$

$$\le 3\frac{n}{\delta} + \frac{n}{2\delta}\log(\delta^2 n)\ , \tag{2.4.8}$$

where the last inequality holds for any sufficiently large n. Combining the above inequality and the definition of i_0, we deduce that

$$(\ s_{t_n(0)} =)\ \mathbf{E}_{s_0}\left[|S_{t_n(0)}|\mathbf{1}_{\{\tau_0 > t_n(0)\}}\right] \le 1/\sqrt{\delta n}\ . \tag{2.4.9}$$

Thus, by Corollary 2.3.6 (after taking expectation), for some fixed $c > 0$

$$\mathbf{P}(\tau_0 > t_n(\gamma)) \le c/\sqrt{\gamma}\ .$$

For the case $\delta^2 n = O(1)$, choose $a = 2$, that is, $b_i = 2^{-(i+2)}$, and define

$$i_1 = \min\{i : b_i \le n^{-1/4} \vee 5\sqrt{|\delta|}\}\ . \tag{2.4.10}$$

Substituting $a = 2$ in (2.4.7), while noting that $\delta > -\frac{1}{25}b_i^2$ for all $i < i_1$, gives

$$\sum_{i=1}^{i_1}(u_{i+1}-u_i) \le \sum_{i=1}^{i_1}\frac{20n}{20\delta + b_i^2} \le 100\sum_{i=1}^{i_1}\frac{n}{b_i^2} \le 200\frac{n}{b_{i_1}^2}$$

$$\le \left(200n^{3/2} \wedge 8\frac{n}{|\delta|}\right) \le 200n^{3/2}\ .$$

where the last inequality in the first line incorporated the geometric sum over $\{b_i^{-2}\}$. By

(2.4.10),

$$b_{i_1} \leq n^{-1/4} \vee 5\sqrt{|\delta|} \leq n^{-1/4}\left(1 + 5(\delta^2 n)^{1/4}\right) ,$$

and as in the subcritical case, we now combine the above results with an application of Corollary 2.3.6 (after taking expectation), and deduce that for some absolute constant $c > 0$,

$$\mathbf{P}(\tau_0 > t_n(\gamma)) \leq c/\sqrt{\gamma} ,$$

as required. ∎

Apart from drifting toward 0, and as we had previously mentioned, the magnetization chain at high temperatures is in fact contracting; this is a special case of the following lemma.

Lemma 2.4.4. *Let (S_t) and $(\tilde{S}_t)$ be the corresponding magnetization chains of two instances of the Glauber dynamics for some $\beta = 1 - \delta$ (where δ is not necessarily positive), and put $D_t := S_t - \tilde{S}_t$. The following then holds:*

$$\mathbf{E}[D_{t+1} - D_t \mid D_t] \leq -\frac{\delta}{n}D_t + \frac{|D_t|}{n^2} + O(n^{-4}) . \tag{2.4.11}$$

Proof. By definition (recall (2.3.5)), we have

$$
\begin{aligned}
\mathbf{E}[D_{t+1} - D_t \mid D_t] &= \mathbf{E}[S_{t+1} - S_t + \tilde{S}_t - \tilde{S}_{t+1} \mid D_t] \\
&= \frac{\tilde{S}_t - S_t}{n} + \left[\varphi(S_t) - \varphi(\tilde{S}_t)\right] - \left[\psi(S_t) - \psi(\tilde{S}_t)\right] .
\end{aligned}
$$

The Mean Value Theorem implies that

$$\varphi(S_t) - \varphi(\tilde{S}_t) \leq \frac{\beta}{n}(S_t - \tilde{S}_t) ,$$

and applying Taylor expansions on $\tanh(x)$ around βS_t and $\beta\tilde{S}_t$, we deduce that

$$\psi(S_t) - \psi(\tilde{S}_t) = \frac{S_t}{n^2\cosh^2(\beta S_t)} - \frac{\tilde{S}_t}{n^2\cosh^2(\beta\tilde{S}_t)} + O\left(\frac{1}{n^4}\right) .$$

Since the derivative of the function $x/\cosh^2(\beta x)$ is bounded by 1, another application of the Mean Value Theorem gives

$$\left|\psi(S_t) - \psi(\tilde{S}_t)\right| \leq \frac{|S_t - \tilde{S}_t|}{n^2} + O\left(\frac{1}{n^4}\right) .$$

Altogether, we obtain (2.4.11), as required. ∎

Indeed, the above lemma ensures that in the high temperature regime, $\beta = 1 - \delta$ where $\delta > 0$, the magnetization chain is contracting:

$$\mathbf{E}\left[|D_{t+1}| \,\middle|\, D_t\right] \leq \left(1 - \frac{\delta}{2n}\right)|D_t| \;\; \text{for any sufficiently large } n \;. \qquad (2.4.12)$$

We are now ready to prove that hitting near 0 essentially ensures the mixing of the magnetization.

Lemma 2.4.5. *Let $\beta = 1 - \delta$ for $\delta > 0$ with $\delta^2 n \to \infty$, (X_t) and $(\tilde{X}_t)$ be two instances of the dynamics started from arbitrary states σ_0 and $\tilde{\sigma}_0$ respectively, and (S_t) and $(\tilde{S}_t)$ be their corresponding magnetization chains. Let τ_{mag} denote the coalescence time $\tau_{\mathrm{mag}} := \min\{t : S_t = \tilde{S}_t\}$, and $t_n(\gamma)$ be as defined in Theorem 2.4.3. Then there exists some constant $c > 0$ such that*

$$\mathbf{P}\left(\tau_{\mathrm{mag}} > t_n(3\gamma)\right) \leq c/\sqrt{\gamma} \;\; \text{for all } \gamma > 0 \;. \qquad (2.4.13)$$

Proof. Set $T = t_n(\gamma)$. We claim that the following holds for large n:

$$|\mathbf{E}S_t| \leq \frac{2}{\sqrt{\delta n}} \;\; \text{and} \;\; |\mathbf{E}\tilde{S}_t| \leq \frac{2}{\sqrt{\delta n}} \;\; \text{for all } t \geq T \;. \qquad (2.4.14)$$

To see this, first consider the case where n is even. The above inequality then follows directly from (2.4.9) and the decreasing property of s_t (see (2.4.6)), combined with the fact that $\mathbf{E}_0 S_t = 0$ (and thus $\mathbf{E}S_t = 0$ for all $t \geq \tau_0$). In fact, in case n is even, $|\mathbf{E}S_t|$ and $|\mathbf{E}\tilde{S}_t|$ are both at most $1/\sqrt{\delta n}$ for all $t \geq T$. For the case of n odd (where there is no 0 state for the magnetization chain, and τ_0 is the hitting time to $\pm\frac{1}{n}$), a simple way to show that (2.4.14) holds is to bound $|\mathbf{E}_{\frac{1}{n}} S_t|$. By definition, $P_M(\frac{1}{n}, \frac{1}{n}) \geq P_M(\frac{1}{n}, -\frac{1}{n})$ (see (2.3.3)). Combined with the symmetry of the positive and negative parts of the magnetization chain, one can then verify by induction that $P_M^t(\frac{1}{n}, \frac{k}{n}) \geq P_M^t(\frac{1}{n}, -\frac{k}{n})$ for any odd $k > 0$ and any t. Therefore, by symmetry as well as the fact that $\mathbf{E}_{s_0} S_t \leq s_0$ for positive s_0, we conclude that $|\mathbf{E}_{\frac{1}{n}} S_t|$ is decreasing with t, and thus is bounded by $\frac{1}{n}$. This implies that (2.4.14) holds for odd n as well.

Combining (2.4.14) with the Cauchy-Schwartz inequality we obtain that for any $t \geq T$

$$\mathbf{E}|S_t - \tilde{S}_t| \leq \mathbf{E}|S_t| + \mathbf{E}|\tilde{S}_t| \leq \sqrt{\mathrm{Var}(S_t) + \frac{4}{\delta n}} + \sqrt{\mathrm{Var}(\tilde{S}_t) + \frac{4}{\delta n}} \;.$$

Now, combining Lemma 2.3.4 and Lemma 2.4.4 (and in particular, (2.4.12)), we deduce that

$\text{Var}\, S_t \leq \frac{4}{\delta n}$, and plugging this into the above inequality gives

$$\mathbf{E}|S_t - \tilde{S}_t| \leq \frac{10}{\sqrt{\delta n}} \quad \text{for any } t \geq T \ .$$

We next wish to show that within $2\gamma n/\delta$ additional steps, S_t and $\tilde{S}_t$ coalesce with probability at least $1 - c/\sqrt{\gamma}$ for some constant $c > 0$.

Consider time T, and let $D_t := S_t - \tilde{S}_t$. Recall that we have already established that

$$\mathbf{E}D_T \leq 10/\sqrt{\delta n} \ , \tag{2.4.15}$$

and assume without loss of generality that $D_T > 0$. We now run the magnetization chains S_t and $\tilde{S}_t$ independently for $T \leq t \leq \tau_1$, where

$$\tau_1 := \min\left\{t \geq T : D_t \in \{0, -\tfrac{2}{n}\}\right\} \ ,$$

and let $\mathcal{F}_t$ be the σ-field generated by these two chains up to time t. By Lemma 2.4.4, we deduce that for sufficiently large values of n, if $D_t > 0$ then

$$\mathbf{E}[D_{t+1} - D_t \mid \mathcal{F}_t] \leq -\frac{\delta}{2n}D_t \leq 0 \ , \tag{2.4.16}$$

and D_t is a supermartingale with respect to $\mathcal{F}_t$. Hence, so is

$$W_t := D_{T+t} \cdot \frac{n}{2}\mathbf{1}_{\{\tau_1 > t\}} \ ,$$

and it is easy to verify that W_t satisfies the conditions of Lemma 2.3.5 (recall the upper bound on the holding probability of the magnetization chain, as well as the fact that at most one spin is updated at any given step). Therefore, for some constant $c > 0$,

$$\begin{aligned}\mathbf{P}\left(\tau_1 > t_n(2\gamma) \mid D_T\right) &= \mathbf{P}(W_0 > 0, W_1 > 0, \ldots, W_{t_n(2\gamma)-T} > 0 \mid D_T) \\ &\leq \frac{cnD_T}{\sqrt{\gamma n/\delta}} \ .\end{aligned}$$

Taking expectation and plugging in (2.4.15), we get that for some constant c',

$$\mathbf{P}\left(\tau_1 > t_n(2\gamma)\right) \leq \frac{c'}{\sqrt{\gamma}} \ . \tag{2.4.17}$$

From time τ_1 and onward, we couple S_t and $\tilde{S}_t$ using a monotone coupling, thus D_t becomes

a non-negative supermartingale with $D_{\tau_1} \leq \frac{2}{n}$. By (2.4.16),

$$\mathbf{E}\left[D_{t+1} - D_t \mid \mathcal{F}_t\right] \leq -\frac{\delta}{n^2} \quad \text{for } \tau_1 \leq t < \tau_{\text{mag}} \;,$$

and therefore, the Optional Stopping Theorem for non-negative supermartingales implies that, for some constant c'',

$$\mathbf{P}\left(\tau_{\text{mag}} - \tau_1 \geq n/\delta\right) \leq \frac{\mathbf{E}(\tau_{\text{mag}} - \tau_1)}{n/\delta} \leq \frac{c''}{\gamma} \;. \tag{2.4.18}$$

Combining (2.4.17) and (2.4.18) we deduce that for some constant c,

$$\mathbf{P}\left(\tau_{\text{mag}} > t_n(3\gamma)\right) \leq \frac{c}{\sqrt{\gamma}} \;,$$

completing the proof. ∎

Lower bound

We need to prove that the following statement holds for the distance of the magnetization at time t from stationarity:

$$\lim_{\gamma \to \infty} \liminf_{n \to \infty} \; d_n \left(\frac{1}{2} \cdot \frac{n}{\delta} \log(\delta^2 n) - \gamma \frac{n}{\delta}\right) = 1 \;. \tag{2.4.19}$$

The idea is to show that, at time $\frac{1}{2} \cdot \frac{n}{\delta} \log(\delta^2 n) - \gamma \frac{n}{\delta}$, the expected magnetization remains large. Standard concentration inequalities will then imply that the magnetization will typically be significantly far from 0, unlike its stationary distribution.

To this end, we shall first analyze the third moment of the magnetization chain. Recalling the transition rule (2.3.3) of S_t under the notations (2.3.1),(2.3.2)

$$p^+(s) = \frac{1 + \tanh(\beta s)}{2} \;, \quad p^-(s) = \frac{1 - \tanh(\beta s)}{2} \;,$$

the following holds:

$$\mathbf{E}\left[S_{t+1}^3 \mid S_t = s\right]$$

$$= \frac{1+s}{2}p^-(s-n^{-1})\left(s-\frac{2}{n}\right)^3 + \frac{1-s}{2}p^+(s+n^{-1})\left(s+\frac{2}{n}\right)^3$$

$$+ \left(1 - \frac{1+s}{2}p^-(s-n^{-1}) - \frac{1-s}{2}p^+(s+n^{-1})\right)s^3$$

$$= s^3 + \frac{6s^2}{n}\cdot\frac{1}{4}\left(-2s + \tanh\left(\beta(s-n^{-1})\right) + \tanh\left(\beta(s+n^{-1})\right)\right.$$

$$+ s\left(\tanh\left(\beta(s-n^{-1})\right) - \tanh\left(\beta(s+n^{-1})\right)\right) + c_1\frac{s}{n^2} + \frac{c_2}{n^3}. \tag{2.4.20}$$

As $\tanh(x) \le x$ for $x \ge 0$, for every $s > 0$ we get

$$\mathbf{E}\left[S_{t+1}^3 \mid S_t = s\right] \le s^3 + \frac{3s^2}{2n}\left(-2s + \beta(s-n^{-1}) + \beta(s+n^{-1})\right) + c_1\frac{s}{n^2} + \frac{c_2}{n^3}$$

$$= s^3 - 3\frac{\delta}{n}s^3 + \frac{c_1}{n^2}s + \frac{c_2}{n^3}. \tag{2.4.21}$$

If $s = 0$, the above also holds, since in that case $|S_{t+1}|^3 \le (2/n)^3$. Finally, by symmetry, if $s < 0$ then the distribution of $|S_{t+1}^3| = -S_{t+1}^3$ given $S_t = s$ is the same as that of S_{t+1}^3 given $S_t = |s|$, and altogether we get:

$$\mathbf{E}\left[|S_{t+1}|^3 \mid S_t = s\right] \le |s|^3 - 3\frac{\delta}{n}|s|^3 + \frac{c_1}{n^2}|s| + \frac{c_2}{n^3}.$$

We deduce that

$$\mathbf{E}|S_{t+1}|^3 \le \mathbf{E}\left(|S_t|^3 - 3\frac{\delta}{n}|S_t|^3 + \frac{c_1}{n^2}|S_t| + \frac{c_2}{n^3}\right)$$

$$\le \left(1 - \frac{3\delta}{n}\right)\mathbf{E}|S_t|^3 + \frac{c_1}{n^2}\mathbf{E}|S_t| + \frac{c_2}{n^3}. \tag{2.4.22}$$

Note that the following statement holds for the first moment of S_t:

$$\mathbf{E}_{s_0}\left[|S_t|\right] \le \sqrt{(\mathbf{E}_{s_0}S_t)^2 + \mathrm{Var}_{s_0}(S_t)}$$

$$\le \sqrt{(s_t)^2 + \frac{16}{\delta n}} \le \left(1 - \frac{\delta}{n}\right)^t |s_0| + \frac{4}{\sqrt{\delta n}}.$$

Hence,

$$\mathbf{E}_{s_0}|S_{t+1}|^3 \leq \left(1 - \frac{3\delta}{n}\right)\mathbf{E}_{s_0}|S_t|^3 + \frac{c_1}{n^2}\left(1 - \frac{\delta}{n}\right)^t|s_0| + \frac{2}{n^2\sqrt{\delta n}} + \frac{c_2}{n^3}$$

$$= \eta^3\mathbf{E}_{s_0}|S_t|^3 + \eta^t\frac{c_1}{n^2}|s_0| + \frac{4}{n^2\sqrt{\delta n}} + \frac{c_2'\delta^2}{n^2} \ ,$$

where $\eta = 1 - \delta/n$, and the extra error term involving c_2' absorbs the change of coefficient of $\mathbf{E}_{s_0}|S_t|^3$ and also the $1/n^3$ term. Iterating, we obtain

$$\mathbf{E}_{s_0}|S_{t+1}|^3 \leq \eta^{3t}|s_0|^3 + \eta^t\frac{c_1}{n^2}|s_0|\sum_{j=0}^t \eta^{2j} + \left(\frac{c_1'}{n^2\sqrt{\delta n}} + \frac{c_2'\delta^2}{n^2}\right)\sum_{j=0}^t \eta^{3j}$$

$$\leq \eta^{3t}|s_0|^3 + \eta^t\frac{c_1}{n^2}\cdot\frac{|s_0|}{1-\eta^2} + \left(\frac{c_1'}{n^2\sqrt{\delta n}} + \frac{c_2'\delta^2}{n^2}\right)\cdot\frac{1}{1-\eta^3}$$

$$\leq \eta^{3t}|s_0|^3 + \eta^t\frac{c_1}{\delta n}|s_0| + \frac{c_1'}{(\delta n)^{3/2}} + \frac{c_2'\delta}{n} \ . \tag{2.4.23}$$

Define $Z_t := |S_t|\eta^{-t}$, whence $Z_0 = |S_0| = |s_0|$. Recalling (2.3.5), and combining the Taylor expansion of $\tanh(x)$ given in (2.4.5) with the fact that $|\psi(s)| = O\left(s/n^2\right)$, we get that for $s > 0$

$$\mathbf{E}\left[|S_{t+1}|\,\big|\,S_t = s\right] \geq \eta s - \frac{s^3}{2n} - \frac{s}{n^2} \ .$$

By symmetry, an analogous statement holds for $s < 0$, and altogether we obtain that

$$\mathbf{E}\left[|S_{t+1}|\,\big|\,S_t\right] \geq \eta|S_t| - \frac{|S_t|^3}{2n} - \frac{|S_t|}{n^2} \ . \tag{2.4.24}$$

Remark. Note that (2.4.24) in fact holds for any temperature, having followed from the basic definition of the transition rule of (S_t), rather than from any special properties that this chain may have in the high temperature regime.

Rearranging the terms and multiplying by $\eta^{-(t+1)}$, we obtain that for any sufficiently large n,

$$\mathbf{E}\left[\left(1 - \frac{2}{n^2}\right)Z_t - Z_{t+1} \mid S_t\right] \leq \frac{1}{n}\eta^{-t}|S_t|^3 \ ,$$

where we used the fact that $\eta^{-1} \leq 2$. Taking expectation and plugging in (2.4.23), we deduce

that

$$\mathbf{E}_{s_0}\left[\left(1 - \frac{2}{n^2}\right)Z_t - Z_{t+1}\right] \leq \frac{1}{n}\left(\eta^{2t}|s_0|^3 + \frac{c_1}{\delta n}|s_0| + \eta^{-t}\left(\frac{c_1'}{(\delta n)^{3/2}} + \frac{c_2'\delta}{n}\right)\right). \qquad (2.4.25)$$

Set

$$\bar{t} = \frac{n}{2\delta}\log(\delta^2 n) - \gamma n/\delta ,$$

and notice that when n is sufficiently large, $\left(1 - \frac{2}{n^2}\right)^{-(t+1)} \leq 2$ for any $t \leq \bar{t}$. Therefore, multiplying (2.4.25) by $(1 - \frac{2}{n^2})^{-(t+1)}$ and summing over gives:

$$\begin{aligned}
|s_0| - 2\mathbf{E}_{s_0}Z_{\bar{t}} &\leq \frac{2|s_0|^3}{n(1 - \eta^2)} + \bar{t}\frac{c_1}{\delta n^2}|s_0| + \frac{2\eta^{-\bar{t}}}{n(1 - \eta)}\left(\frac{c_1'}{(\delta n)^{3/2}} + \frac{c_2'\delta}{n}\right) \\
&\leq \frac{2|s_0|^3}{\delta} + \frac{c_1\log(\delta^2 n)}{2\delta^2 n}|s_0| + \frac{c_1'}{\delta^{3/2}n} + \frac{c_2}{\delta n^{3/2}} + \frac{c_2'\delta}{\sqrt{n}} \\
&= \frac{2|s_0|^3}{\delta} + o(\sqrt{\delta} + |s_0|) ,
\end{aligned}$$

where the last inequality follows from the assumption $\delta^2 n \to \infty$. We now select $s_0 = \sqrt{\delta}/3$, which gives

$$\sqrt{\delta}/3 - 2\mathbf{E}_{s_0}Z_{\bar{t}} \leq 2\sqrt{\delta}/27 + o(\sqrt{\delta}) ,$$

and for a sufficiently large n we get

$$\mathbf{E}_{s_0}Z_{\bar{t}} \geq \sqrt{\delta}/9 .$$

Recalling the definition of Z_t, and using the well known fact that $(1 - x) \geq \exp(-x/(1 - x))$ for $0 < x < 1$, we get that for a sufficiently large n,

$$\mathbf{E}_{s_0}|S_{\bar{t}}| \geq \eta^{\bar{t}}\sqrt{\delta}/9 \geq \frac{\mathrm{e}^{\gamma/2}}{10\sqrt{\delta n}} =: L . \qquad (2.4.26)$$

Lemma 2.3.4 implies that $\max\{\text{Var}_{s_0}(S_t), \text{Var}_{\mu_n}(\tilde{S}_t)\} \leq 16/\delta n$. Therefore, recalling that $\mathbf{E}_{\mu_n}\tilde{S}_{\bar{t}} = 0$, Chebyshev's inequality gives

$$\mathbf{P}_{s_0}(|S_{\bar{t}}| \leq L/2) \leq \mathbf{P}_{s_0}(||S_{\bar{t}}| - \mathbf{E}_{s_0}|S_{\bar{t}}|| \geq L/2) \leq \frac{16/(\delta n)}{L^2/4} = c\mathrm{e}^{-\gamma} ,$$

$$\mathbf{P}_{\mu_n}(|\tilde{S}_{\bar{t}}| \geq L/2) \leq \frac{16/(\delta n)}{L^2/4} = c\mathrm{e}^{-\gamma} .$$

Hence, letting π denote the stationary distribution of S_t, and A_L denote the set $\left[-\frac{L}{2}, \frac{L}{2}\right]$, we obtain that

$$\|\mathbf{P}_{s_0}(S_{\bar{t}} \in \cdot) - \pi\|_{\mathrm{TV}} \geq \pi(A_L) - \mathbf{P}_{s_0}(|S_{\bar{t}}| \in A_L) \geq 1 - 2\mathrm{ce}^{-\gamma} ,$$

which immediately gives (2.4.19). ∎

2.4.2 Full Mixing of the Glauber dynamics

In order to boost the mixing of the magnetization into the full mixing of the configurations, we will need the following result, which was implicitly proved in [53, Sections 3.3, 3.4] using a Two Coordinate Chain analysis. Although the authors of [53] were considering the case of $0 < \beta < 1$ fixed, one can follow the same arguments and extend this result to any $\beta < 1$. Following is this generalization of their result:

Theorem 2.4.6 ([53]). *Let (X_t) be an instance of the Glauber dynamics and μ_n the stationary distribution of the dynamics. Suppose X_0 is supported by*

$$\Omega_0 := \{\sigma \in \Omega : |S(\sigma)| \leq 1/2\} .$$

For any $\sigma_0 \in \Omega_0$ and $\tilde{\sigma} \in \Omega$, we consider the dynamics (X_t) starting from σ_0 and an additional Glauber dynamics $(\tilde{X}_t)$ starting from $\tilde{\sigma}$, and define:

$$\tau_{\mathrm{mag}} := \min\{t : S(X_t) = S(\tilde{X}_t)\} ,$$
$$U(\sigma) := |\{i : \sigma(i) = \sigma_0(i) = 1\}| , \quad V(\sigma) := |\{i : \sigma(i) = \sigma_0(i) = -1\}| ,$$
$$\Xi := \{\sigma : \min\{U(\sigma), U(\sigma_0) - U(\sigma), V(\sigma), V(\sigma_0) - V(\sigma))\} \geq n/20\} ,$$
$$R(t) := \left|U(X_t) - U(\tilde{X}_t)\right| ,$$
$$H_1(t) := \{\tau_{\mathrm{mag}} \leq t\} , \quad H_2(t_1, t_2) := \cap_{i=t_1}^{t_2}\{X_i \in \Xi \wedge \tilde{X}_i \in \Xi\} .$$

Then for any possible coupling of X_t and $\tilde{X}_t$, the following holds:

$$\max_{\sigma_0 \in \Omega_0}\|\mathbf{P}_{\sigma_0}(X_{r_2} \in \cdot) - \mu_n\|_{\mathrm{TV}} \leq \max_{\substack{\sigma_0 \in \Omega_0 \\ \tilde{\sigma} \in \Omega}} \Bigg[\mathbf{P}_{\sigma_0, \tilde{\sigma}}(\overline{H_1(r_1)})$$

$$+ \mathbf{P}_{\sigma_0, \tilde{\sigma}}(R_{r_1} > \alpha\sqrt{n/\delta}) + \mathbf{P}_{\sigma_0, \tilde{\sigma}}(\overline{H_2(r_1, r_2)}) + \frac{\alpha c_1}{\sqrt{r_2 - r_1}} \cdot \sqrt{\frac{n}{\delta}}\Bigg] , \qquad (2.4.27)$$

where $r_1 < r_2$ and $\alpha > 0$.

The rest of this subsection will be devoted to establishing a series of properties satisfied

by the magnetization throughout the mildly subcritical case, in order to ultimately apply the above theorem.

First, we shall show that any instance of the Glauber dynamics concentrates on Ω_0 once it performs an initial burn-in period of n/δ steps. It suffices to show this for the dynamics started from $s_0 = 1$: to see this, consider a monotone-coupling of the dynamics (X_t) starting from an arbitrary configuration, together with two additional instances of the dynamics, (X_t^+) starting from $s_0 = 1$ (from above) and (X_t^-) starting from $s_0 = -1$ (from below). By definition of the monotone-coupling, the chains (X_t^+) and (X_t^-) "trap" the chain (X_t), and by symmetry it indeed remains to show that

$$\mathbf{P}_1(|S_{t_0}| \leq 1/2) = 1 - o(1) \ , \ \text{where } t_0 = n/\delta \ .$$

Recalling (2.4.4), we have $s_{t+1} \leq (1 - \frac{\delta}{n})s_t$ where $s_t = \mathbf{E}\left[|S_t|\mathbf{1}_{\{\tau_0 > t\}}\right]$, thus

$$\mathbf{E}_1\left[|S_{t_0}|\mathbf{1}_{\{\tau_0 > t_0\}}\right] \leq \mathrm{e}^{-1} \ .$$

Adding this to the fact that $\mathbf{E}_1 S_{t_0}\mathbf{1}\{\tau_0 \leq t_0\} = 0$, which follows immediately from symmetry, we conclude that $\mathbf{E}_1 S_{t_0} \leq \mathrm{e}^{-1}$. Next, applying Lemma 2.3.4 to our case and noting that (2.3.7) holds for $\rho' = 1 - \frac{1}{n}\left(1 - n\tanh(\frac{\beta}{n})\right) \leq 1 - \frac{\delta}{n}$, we conclude that

$$\mathrm{Var}(S_t) \leq \nu_1 \frac{n}{\delta} \leq \left(\frac{4}{n}\right)^2 \frac{n}{\delta} = \frac{16}{\delta n} \ \text{ for all } t \ .$$

Hence, Chebyshev's inequality gives that $|S_{t_0}| \leq 1/2$ with high probability. We may therefore assume henceforth that our initial configuration already belongs to some good state $\sigma_0 \in \Omega_0$.

Next, set:

$$T := t_n(\gamma) \ , r_0 := t_n(2\gamma) \ , r_1 := t_n(3\gamma) \ , r_2 := t_n(4\gamma) \ .$$

We will next bound the terms in the righthand side of (2.4.27) in order. First, recall that Lemma 2.4.5 already provided us with a bound on the probability of $\overline{H_1(r_1)}$, by stating there for constant $c > 0$

$$\mathbf{P}(\tau_{\mathrm{mag}} > r_1) \leq \frac{c}{\sqrt{\gamma}} \ . \tag{2.4.28}$$

Our next task is to provide an upper bound on R_{r_1}, and namely, to show that it typically has order at most $\sqrt{n/\delta}$. In order to obtain such a bound, we will analyze the sum of the

spins over the set $B := \{i : \sigma_0(i) = 1\}$. Define

$$M_t(B) := \frac{1}{2} \sum_{i \in B} X_t(i) \ ,$$

and consider the monotone-coupling of (X_t) with the chains (X_t^+) and (X_t^-) starting from the all-plus and all-minus positions respectively, such that $X_t^- \le X_t \le X_t^+$. By defining M_t^+ and M_t^- accordingly, we get that

$$\mathbf{E}(M_t(B))^2 \le \mathbf{E}(M_t^+(B))^2 + \mathbf{E}(M_t^-(B))^2 = 2\mathbf{E}(M_t^+(B))^2 \ .$$

By (2.4.14), we immediately get that for $t \ge T$, $|\mathbf{E}M_t^+(B)| \le \sqrt{\frac{n}{\delta}}$. We will next bound the variance of $M_t^+(B)$, by considering the following two cases:

(i) If every pair of spins of X_t^+ is positively correlated (since X_0^+ is the all-plus configuration, by symmetry, the covariances of each pair of spins is the same), then we can infer that

$$\operatorname{Var}(M_t^+(B)) \le \operatorname{Var}\left(\frac{1}{2} \sum_{i \in [n]} X_t^+(i)\right) = \frac{n^2}{4} \operatorname{Var}\left(S(X_t^+)\right) \le \frac{4n}{\delta} \ .$$

(ii) Otherwise, every pair of spins of X_t^+ is negatively correlated, and it follows that

$$\operatorname{Var}(M_t^+(B)) \le \sum_{i \in B} \operatorname{Var}\left(\frac{1}{2} X_t^+(i)\right) \le \frac{n}{4} \ .$$

Altogether, we conclude that for all $t \ge T$,

$$\mathbf{E}|M_t(B)| \le \sqrt{\mathbf{E}\left(M_t(B)\right)^2} \le \sqrt{2\operatorname{Var}(M_t^+(B) + 2\left(\mathbf{E}M_t(B)\right)^2}$$

$$\le \sqrt{\frac{8n}{\delta} + \frac{2n}{\delta}} \le 8\sqrt{\frac{n}{\delta}} \ . \tag{2.4.29}$$

This immediately implies that

$$\mathbf{E}R_{r_1} = \mathbf{E}|M_{r_1}(B) - \tilde{M}_{r_1}(B)| \le \mathbf{E}|M_{r_1}(B)| + \mathbf{E}|\tilde{M}_{r_1}(B)| \le 16\sqrt{\frac{n}{\delta}} \ ,$$

and an application of Markov's inequality now gives

$$\mathbf{P}(R_{r_1} \ge \alpha\sqrt{\frac{n}{\delta}}) \le \frac{16}{\alpha} \ . \tag{2.4.30}$$

It remains to bound the probability of $\overline{H_2(r_1, r_2)}$. Define:

$$Y := \sum_{r_1 \leq t \leq r_2} \mathbf{1}\{|M_t(B)| > n/64\} \, ,$$

and notice that

$$\mathbf{P}\left(\bigcup_{t=r_1}^{r_2} \{|M_t(B)| \geq n/32\} \right) \leq \mathbf{P}(Y > n/64) \leq \frac{c_0 \mathbf{E}[Y]}{n} \, .$$

Recall that the second inequality of (2.4.29) actually gives $\mathbf{E}|M_t(B)|^2 \leq \frac{5n}{\delta}$. Hence, a standard second moment argument gives

$$\mathbf{P}(|M_t(B)| > n/64) = O\left(\frac{1}{\delta n} \right) \, .$$

Altogether, $\mathbf{E}_{\sigma_0} Y = O(1/\delta^2)$ and

$$\mathbf{P}_{\sigma_0}\left(\bigcup_{t=r_1}^{r_2} \{|M_t(B)| \geq n/32\} \right) = O\left(\frac{1}{\delta^2 n} \right) \, .$$

Applying an analogous argument to the chain $(\tilde{X}_t)$, we obtain that

$$\mathbf{P}_{\tilde{\sigma}}\left(\bigcup_{t=r_1}^{r_2} \left\{ |\tilde{M}_t(B)| \geq n/32 \right\} \right) = O\left(\frac{1}{\delta^2 n} \right) \, ,$$

and combining the last two inequalities, we conclude that

$$\mathbf{P}_{\sigma_0, \tilde{\sigma}}\left(\overline{H_2(r_1, r_2)} \right) = O\left(\frac{1}{\delta^2 n} \right) \, . \tag{2.4.31}$$

Finally, we have established all the properties needed in order to apply Theorem 2.4.6. At the cost of a negligible number of burn-in steps, the state of (X_t) with high probability belongs to Ω_0. We may thus plug in (2.4.28), (2.4.30) and (2.4.31) into Theorem 2.4.6, choosing $\alpha = \sqrt{\gamma}$, to obtain (2.4.1).

2.4.3   Spectral gap Analysis

By Proposition 2.3.9, it suffices to determine the spectral gap of the magnetization chain. The lower bound will follow from the next lemma of [17] (see also [54, Theorem 13.1]) along

with the contraction properties of the magnetization chain.

Lemma 2.4.7 ([17])**.** *Suppose Ω is a metric space with distance ρ. Let P be a transition matrix for a Markov chain, not necessarily reversible. Suppose there exists a constant $\theta < 1$ and for each $x, y \in \Omega$, there is a coupling (X_1, Y_1) of $P(x, \cdot)$ and $P(y, \cdot)$ satisfying*

$$\mathbf{E}_{x,y}(\rho(X_1, Y_1)) \leq \theta \rho(x, y) .$$

If λ is an eigenvalue of P different from 1, then $|\lambda| \leq \theta$. In particular, the spectral gap satisfies $\mathtt{gap} \geq 1 - \theta$.

Recalling (2.4.12), the monotone coupling of S_t and $\tilde{S}_t$ implies that

$$\mathbf{E}_{s,\tilde{s}}\big|S_1 - \tilde{S}_1\big| \leq \left(1 - \frac{\delta}{n} + o\Big(\frac{\delta}{n}\Big)\right) |s - \tilde{s}| .$$

Therefore, Lemma 3.4.8 ensures that $\mathtt{gap} \geq (1 + o(1))\frac{\delta}{n}$.

It remains to show a matching upper bound on $\mathtt{gap}$, the spectral gap of the magnetization chain. Let M be the transition kernel of this chain, and π be its stationary distribution. Similar to our final argument in Proposition 2.3.9 (recall (3.2.1)), we apply the Dirichlet representation for the spectral gap (as given in [54, Lemma 13.7]) with respect to the function f being the identity map $\mathbf{1}$ on the space of normalized magnetization, we obtain that

$$\mathtt{gap} \leq \frac{\mathcal{E}(\mathbf{1})}{\langle \mathbf{1}, \mathbf{1} \rangle_\pi} = \frac{\langle (I - M)\mathbf{1}, \mathbf{1} \rangle_\pi}{\langle \mathbf{1}, \mathbf{1} \rangle_\pi} = 1 - \frac{\mathbf{E}_\pi \left[\mathbf{E}\left[S_t S_{t+1} \mid S_t\right]\right]}{\mathbf{E}_\pi S_t^2} , \qquad (2.4.32)$$

where $\mathbf{E}_\pi S_t^k$ is the k-th moment of the stationary magnetization chain (S_t). Recall (2.4.24) (where $\eta = 1 - \frac{\delta}{n}$), and notice that the following slightly stronger inequality in fact holds:

$$\mathbf{E}\left[\mathrm{sign}(S_t)S_{t+1} \mid S_t\right] \geq \eta|S_t| - \frac{|S_t|^3}{2n} - \frac{|S_t|}{n^2} .$$

(to see this, one needs to apply the same argument that led to (2.4.24), then verify the special cases $S_t \in \{0, \frac{1}{n}\}$). It thus follows that

$$\mathbf{E}\left[S_t S_{t+1} \mid S_t\right] \geq \eta S_t^2 - \frac{S_t^4}{2n} - \frac{S_t^2}{n^2} ,$$

and plugging the above into (2.4.32) we get

$$\mathtt{gap} \leq \frac{\delta}{n} + \frac{1}{2n} \cdot \frac{\mathbf{E}_\pi S_t^4}{\mathbf{E}_\pi S_t^2} + \frac{1}{n^2} . \qquad (2.4.33)$$

In order to bound the second term in (2.4.33), we need to give an upper bound for the fourth moment in terms of the second moment. The next argument is similar to the one used earlier to bound the third moment of the magnetization chain (see (2.4.20)), and hence will be described in a more concise manner.

For convenience, we use the abbreviations $h^+ := \tanh\left(\beta(s + n^{-1})\right)$ and $h^- := \tanh\left(\beta(s - n^{-1})\right)$. By definition (see (2.3.3)) the following then holds:

$$
\begin{aligned}
\mathbf{E}[S_{t+1}^4 \mid S_t = s] = s^4 &+ \frac{2}{n}s^3\left(-2s + h^- + h^+ + sh^- - h^+\right) \\
&+ \frac{6}{n^2}s^2\left(2 + h^+ - h^- - sh^- + h^+\right) \\
&+ \frac{8}{n^3}s^3\left(-2s + h^- + h^+ + sh^- - h^+\right) \\
&+ \frac{4}{n^4}\left(2 + h^+ - h^- - sh^- + h^+\right) \\
&\leq \left(1 - \frac{4\delta}{n}\right)s^4 + \frac{12}{n^2}s^2 + \frac{16}{n^4} .
\end{aligned}
$$

Now, taking expectation and letting the S_t be distributed according to π, we obtain that

$$
\mathbf{E}_\pi S_t^4 \leq \frac{3}{\delta n}\mathbf{E}_\pi S_t^2 + \frac{4}{\delta n^3} .
$$

Recalling that, as the spins are positively correlated, $\mathrm{Var}_\pi(S_t) \geq \frac{1}{n}$, we get

$$
\mathbf{E}_\pi S_t^4 \leq \left(3 + \frac{4}{n}\right)\frac{\mathbf{E}_\pi S_t^2}{n\delta} . \tag{2.4.34}
$$

Plugging (2.4.34) into (2.4.33), we conclude that

$$
\mathtt{gap} \leq \frac{\delta}{n}\left(1 + O\left(\frac{1}{\delta^2 n}\right)\right) = (1 + o(1))\frac{\delta}{n} .
$$

2.5   The critical window

In this section we prove Theorem 4, which establishes that the critical window has a mixing-time of order $n^{3/2}$ without a cutoff, as well as a spectral-gap of order $n^{-3/2}$.

2.5.1 Upper bound

Let (X_t) denote the Glauber dynamics, started from an arbitrary configuration σ, and let $(\tilde{X}_t)$ denote the dynamics started from the stationary distribution μ_n. As usual, let (S_t) and $(\tilde{S}_t)$ denote the (normalized) magnetization chains of (X_t) and $(\tilde{X}_t)$ respectively.

Let $\varepsilon > 0$. The case $\delta^2 n = O(1)$ of Theorem 2.4.3 implies that, for a sufficiently large $\gamma > 0$, $\mathbf{P}_\sigma \left(\tau_0 \geq \gamma n^{3/2} \right) < \varepsilon$. Plugging this into Lemma 2.3.1, we deduce that there exists some $c_\varepsilon > 0$, such that the chains S_t and $\tilde{S}_t$ coalesce after at most $c_\varepsilon n^{3/2}$ steps with probability at least $1 - \varepsilon$.

At this point, Lemma 2.3.3 implies that (X_t) and $(\tilde{X}_t)$ coalesce after at most $O(n^{3/2}) + O(n \log n) = O(n^{3/2})$ additional steps with probability arbitrarily close to 1, as required.

2.5.2 Lower bound

Throughout this argument, recall that δ is possibly negative, yet satisfies $\delta^2 n = O(1)$. By (2.4.22),

$$
\begin{aligned}
\mathbf{E}|S_{t+1}|^3 &\leq \mathbf{E}\left(|S_t|^3 - 3\frac{\delta}{n}|S_t|^3 + \frac{c_1}{n^2}|S_t| + \frac{c_2}{n^3} \right) \\
&\leq \left(1 - \frac{3\delta}{n} \right) \mathbf{E}|S_t|^3 + \frac{c_1}{n^2}\mathbf{E}|S_t| + \frac{c_2}{n^3} \ .
\end{aligned}
$$

Recalling Lemma 2.4.4, and plugging the fact that $\delta = O(n^{-1/2})$ in (2.4.11), the following holds. If S_t and $\tilde{S}_t$ are the magnetization chains corresponding to two instances of the Glauber dynamics, then for some constant $c > 0$ and any sufficiently large n,

$$
\mathbf{E}_{s,\tilde{s}}|S_1 - \tilde{S}_1| \leq (1 + cn^{-3/2})|s - \tilde{s}| \ . \tag{2.5.1}
$$

Combining this with the extended form of Lemma 2.3.4, as given in (2.3.8), we deduce that if $t \leq \varepsilon n^{3/2}$ for some small fixed $\varepsilon > 0$, then $\mathrm{Var}_{s_0} S_t \leq 4t/n^2$. Therefore,

$$
\mathbf{E}_{s_0}\left[|S_t| \right] \leq \sqrt{|\mathbf{E}_{s_0} S_t|^2 + \mathrm{Var}_{s_0} S_t} \leq \left(1 - \frac{\delta}{n} \right)^t |s_0| + \frac{2\sqrt{t}}{n} \ .
$$

Therefore,

$$\mathbf{E}_{s_0}|S_{t+1}|^3 \leq \left(1 - \frac{3\delta}{n}\right)\mathbf{E}_{s_0}|S_t|^3 + \frac{c_1}{n^2}\left(1 - \frac{\delta}{n}\right)^t|s_0| + \frac{c_1'\sqrt{t}}{n^3}$$

$$\leq \eta^3\mathbf{E}_{s_0}|S_t|^3 + \eta^t\frac{c_1}{n^2}|s_0| + \frac{c_1'\sqrt{t}}{n^3} \ ,$$

where again $\eta = 1 - \delta/n$. Iterating, we obtain

$$\mathbf{E}_{s_0}|S_{t+1}|^3 \leq \eta^{3t}|s_0|^3 + \eta^t\frac{c_1}{n^2}|s_0|\sum_{j=0}^{t}\eta^{2j} + \frac{c_1'\sqrt{t}}{n^3}\sum_{j=0}^{t}\eta^{3j}$$

$$\leq \eta^{3t}|s_0|^3 + \eta^t\frac{c_1}{n^2}\cdot\frac{\eta^{2t-1}-1}{\eta^2-1}|s_0| + \frac{c_1'\sqrt{t}}{n^3}\cdot\frac{\eta^{3t}-1}{\eta^3-1}$$

$$\leq \eta^{3t}|s_0|^3 + \eta^t\frac{c_1}{n^2}\cdot 2t|s_0| + \frac{c_1'\sqrt{t}}{n^3}\cdot 3t \ , \tag{2.5.2}$$

where the last inequality holds for sufficiently large n and $t \leq \varepsilon n^{3/2}$ with $\varepsilon > 0$ small enough (such a choice ensures that η^t will be suitably small). Define $Z_t := |S_t|\eta^{-t}$, whence $Z_0 = |S_0| = |s_0|$. Applying (2.4.24) (recall that it holds for any temperature) and using the fact that $\eta^{-1} \leq 2$, we get

$$\mathbf{E}[Z_{t+1} \mid S_t] \geq Z_t - \frac{1}{n}\left(\eta^{-t}|S_t|^3 + O(1/n)\right),$$

for n large enough, hence

$$\mathbf{E}[Z_t - Z_{t+1} \mid S_t] \leq \frac{1}{n}\left(\eta^{-t}|S_t|^3 + O(1/n)\right) \ .$$

Taking expectation and plugging in (2.4.23),

$$\mathbf{E}_{s_0}[Z_t - Z_{t+1}] \leq \frac{1}{n}\left(\eta^{2t}|s_0|^3 + \frac{c_2t}{n^2}|s_0| + \eta^{-t}\frac{c_2't^{3/2}}{n^3} + O(1/n)\right) \ . \tag{2.5.3}$$

Set $\bar{t} = n^{3/2}/A^4$ for some large constant A such that $\frac{1}{2} \leq \eta^{\bar{t}} \leq 2$. Summing over (2.5.3) we obtain that

$$|s_0| - \mathbf{E}_{s_0} Z_{\bar{t}} \leq \frac{1 - \eta^{2\bar{t}}}{n(1 - \eta^2)}|s_0|^3 + \bar{t}^2 \frac{c_2}{n^3}|s_0| + 2\eta^{-\bar{t}} \cdot \bar{t}^{5/2}/n^4 + O(\bar{t}/n^2)$$

$$\leq \frac{2}{A^4}\sqrt{n}|s_0|^3 + \frac{c_2}{A^8}|s_0| + \frac{2}{A^{10}}e^{\sqrt{\delta^2 n}/A^4}n^{-1/4} + O(n^{-1/2}) .$$

We now select $s_0 = An^{-1/4}$ for some large constant A; this gives

$$An^{-1/4} - \mathbf{E}_{s_0} Z_{\bar{t}} \leq \left(\frac{2}{A} + \frac{c_2}{A^7} + \frac{2}{A^{10}}e^{\sqrt{\delta^2 n}/A^4}\right) n^{-1/4} + O(n^{-1/2}) .$$

Choosing A large enough to swallow the constant c_2 as well as the term $\delta^2 n$ (using the fact that $\delta^2 n$ is bounded), we obtain that

$$\mathbf{E}_{s_0} Z_{\bar{t}} \geq \frac{1}{2}An^{-1/4} .$$

Translating Z_t back to $|S_t|$, we obtain

$$\mathbf{E}_{s_0}|S_{\bar{t}}| \geq \eta^{\bar{t}} \cdot \frac{1}{2}An^{-1/4} \geq \sqrt{A}n^{-1/4} =: B , \qquad (2.5.4)$$

provided that A is sufficiently large (once again, using the fact that $\eta^{\bar{t}}$ is bounded, this time from below). Since

$$\mathrm{Var}_{s_0}(S_{\bar{t}}) \leq 16\bar{t}/n^2 = \frac{16}{A^4}n^{-1/2} , \qquad (2.5.5)$$

the following concentration result on the stationary chain $(\tilde{S}_t)$ will complete the proof:

$$\mathbf{P}_{\mu_n}(|\tilde{S}_t| \geq An^{-1/4})\} \leq \varepsilon(A) , \text{ and } \lim_{A \to \infty} \varepsilon(A) = 0 . \qquad (2.5.6)$$

Indeed, combining the above two statements, Chebyshev's inequality implies that

$$\|\mathbf{P}_{s_0}(S_{\bar{t}} \in \cdot) - \pi\|_{\mathrm{TV}} \geq \pi([-B/2, B/2]) - \mathbf{P}_{s_0}(|S_{\bar{t}}| \leq B/2)$$

$$\geq 1 - \frac{64}{A^5} - \varepsilon(\sqrt{A}) . \qquad (2.5.7)$$

It remains to prove (2.5.6). Since we are proving a lower bound for the mixing-time, it suffices to consider a sub-sequence of the δ_n-s such that $\delta_n\sqrt{n}$ converges to some constant (possibly 0). The following result establishes the limiting stationary distribution of the magnetization

chain in this case.

Theorem 2.5.1. *Suppose that $\lim_{n\to\infty} \delta_n \sqrt{n} = \alpha \in \mathbb{R}$. The following holds:*

$$\frac{S_{\mu_n}}{n^{-1/4}} \to \exp\left(-\frac{s^4}{12} - \alpha\frac{s^2}{2}\right) . \tag{2.5.8}$$

Proof. We need the following theorem:

Theorem 2.5.2 ([33, Theorem 3.9]). *Let ρ denote some probability measure, and let $S_n(\rho) = \frac{1}{n}\sum_{j=1}^{n} X_j(\rho)$, where the $\{X_j(\rho) : j \in [n]\}$ have joint distribution*

$$\frac{1}{Z_n} \exp\left[\frac{(x_1 + \ldots + x_n)^2}{2n}\right] \prod_{j=1}^{n} d\rho(x_j) ,$$

and Z_n is a normalization constant. Suppose that $\{\rho_n : n = 1, 2, \ldots\}$ are measures satisfying

$$\exp(x^2/2)d\rho_n \to \exp(x^2/2)d\rho . \tag{2.5.9}$$

Suppose further that ρ has the following properties:

1. Pure: *the function*

$$G_\rho(s) := \frac{s^2}{2} - \log \int e^{sx} d\rho(x)$$

   has a unique global minimum.

2. Centered at m: *let m denote the location of the above global minimum.*

3. Strength δ and type k: *the parameters $k, \delta > 0$ are such that*

$$G_\rho(s) = G_\rho(m) + \delta\frac{(s-m)^{2k}}{(2k)!} + o((s-m)^{2k}) ,$$

   where the $o(\cdot)$-term tends to 0 as $s \to m$.

If, for some real numbers $\alpha_1, \ldots, \alpha_{2k-1}$ we have

$$G_{\rho_n}^{(j)}(m) = \frac{\alpha_j}{n^{1-j/2k}} + o(n^{-1+j/2k}) , \quad j = 1, 2, \ldots, 2k - 1 , n \to \infty ,$$

then the following holds:

$$S_n(\rho_n) \to \mathbf{1}_{\{s \neq m\}}$$

and

$$\frac{S_n(\rho_n) - m}{n^{-1/2k}} \to \begin{cases} N\left(-\frac{\alpha_1}{\delta}, \frac{1}{\delta} - 1\right), & \text{if } k = 1 , \\ \exp\left(-\delta\frac{s^2 k}{(2k)!} - \sum_{j=1}^{2k-1} \alpha_j \frac{s^j}{j!}\right), & \text{if } k \geq 2 . \end{cases},$$

where $\delta^{-1} - 1 > 0$ for $k = 1$.

Let ρ denote the two-point uniform measure on $\{-1, 1\}$, and let ρ_n denote the two-point uniform measure on $\{-\beta_n, \beta_n\}$. As $|1 - \beta_n| = \delta_n = O(1/\sqrt{n})$, the convergence requirement (2.5.9) of the measures ρ_n is clearly satisfied. We proceed to verify the properties of ρ:

$$G_\rho(s) = \frac{s^2}{2} - \log \int e^{sx} d\rho(x) = \frac{s^2}{2} - \log \cosh(s) = \frac{s^4}{12} - \frac{s^6}{45} + O(s^8) .$$

This implies that G_ρ has a unique global minimum at $m = 0$, type $k = 2$ and strength $\delta = 2$. As $\delta_n \sqrt{n} \to \alpha$, we deduce that the G_{ρ_n}-s satisfy

$$G_{\rho_n}(s) = \frac{s^2}{2} - \log \cosh(\beta_n s) ,$$
$$G_{\rho_n}^{(1)}(0) = G_{\rho_n}^{(3)}(0) = 0 ,$$
$$G_{\rho_n}^{(2)}(0) = 1 - \beta_n^2 = \delta_n(2 - \delta_n) = \frac{2\alpha}{\sqrt{n}} + o(n^{-1/2}) .$$

This completes the verification of the conditions of the theorem, and we obtain that

$$\frac{S_n(\rho_n)}{n^{-1/4}} \to \exp\left(-\frac{s^4}{12} - \alpha\frac{s^2}{2}\right) . \tag{2.5.10}$$

Recalling that, if $x_i = \pm 1$ is the i-th spin,

$$\mu_n(x_1, \ldots, x_n) = \frac{1}{Z(\beta)} \exp\left(\frac{\beta}{n} \sum_{1 \leq i < j \leq n} x_i x_j\right), \tag{2.5.11}$$

clearly $S_n(\rho_n)$ has the same distribution as S_{μ_n} for any n. This completes the proof of the theorem. ∎

Remark. One can verify that the above analysis of the mixing time in the critical window holds also for the censored dynamics (where the magnetization is restricted to be non-negative, by flipping all spins whenever it becomes negative). Indeed, the upper bound immediately holds as the censored dynamics is a function of the original Glauber dynamics.

For the lower bound, notice that our argument tracked the absolute value of the magnetization chain, and hence can readily be applied to the censored case as-well. Altogether, the censored dynamics has a mixing time of order $n^{3/2}$ in the critical window $1 \pm \delta$ for $\delta = O(1/\sqrt{n})$.

2.5.3 Spectral gap analysis

The spectral gap bound in the critical temperature regime is obtained by combining the above analysis with results of [25] on birth-and-death chains.

The lower bound on `gap` is a direct consequence of the fact that the mixing time has order $n^{3/2}$, and that the inequality $t_{\mathrm{REL}} \leq t_{\mathrm{MIX}}\left(\frac{1}{4}\right)$ always holds. It remains to prove the matching bound $t_{\mathrm{MIX}}\left(\frac{1}{4}\right) = O(t_{\mathrm{REL}})$. Suppose that this is false, that is, $t_{\mathrm{REL}} = o\left(t_{\mathrm{MIX}}\left(\frac{1}{4}\right)\right)$.

Let A be some large constant, and let $s_0 = An^{-1/4}$. Notice that the case $\delta^2 n = O(1)$ in Theorem 2.4.3 implies that $\mathbf{E}_1 \tau_0 = O(n^{3/2})$. Furthermore, by Theorem 2.5.2, there exists a strictly positive function of A, $\varepsilon(A)$, such that $\lim_{A \to \infty} \varepsilon(A) = 0$ and

$$\frac{1}{2}\varepsilon(A) \leq \pi(S \geq s_0) \leq 2\varepsilon(A)$$

for sufficiently large n. Applying Lemma 2.3.7 with $\alpha = \pi(S \geq s_0)$ and $\beta = \frac{1}{2}$ gives $\mathbf{E}_{s_0} \tau_0 = o(n^{3/2})$. As in Subsection 2.5.2, set $\bar{t} = n^{3/2}/A^4$ for some large constant A. Combining Lemma 2.3.8 with Markov's inequality gives the following total variation bound for this birth-and-death chain:

$$\|\mathbf{P}_{s_0}(S_{\bar{t}} \in \cdot) - \pi\|_{\mathrm{TV}} \leq 4\varepsilon(A) + o(1) \ . \tag{2.5.12}$$

However, the lower bound (2.5.7) obtained in Subsection 2.5.2 implies that:

$$\|\mathbf{P}_{s_0}(S_{\bar{t}} \in \cdot) - \pi\|_{\mathrm{TV}} \geq 1 - 4\varepsilon(\sqrt{A}/2) - 64/A^5 \ . \tag{2.5.13}$$

Choosing a sufficiently large constant A, (2.5.12) and (2.5.13) together lead to a contradiction for large n. We conclude that `gap` $= O(n^{-3/2})$, completing the proof.

Note that, as the condition `gap` $\cdot t_{\mathrm{MIX}}(\frac{1}{4}) \to \infty$ is necessary for cutoff in any family of ergodic reversible finite Markov chains (see, e.g., [25]), we immediately deduce that there is no cutoff in this regime.

Remark. It is worth noting that the order of the spectral gap at $\beta_c = 1$ follows from a simpler argument. Indeed, in that case, the *upper bound* on `gap` can alternatively be derived from its Dirichlet representation, similar to the argument that appeared in the proof of Proposition 2.3.9 (where we substitute the identity function, i.e., the sum of spins in the Dirichlet form).

For this argument, one needs a lower bound for the variance of the stationary magnetization. Such a bound is known for $\beta_c = 1$ (see [34]), rather than throughout the critical window.

2.6 Low temperature regime

In this section we prove Theorem 5, which establishes the order of the mixing time and the spectral gap in the super critical regime (where the mixing of the dynamics is exponentially slow and there is no cutoff).

2.6.1 Exponential mixing

Recall that the normalized magnetization chain S_t is a birth-and-death chain on the space $\mathcal{X} = \{-1, -1 + \frac{2}{n}, \ldots, 1 - \frac{2}{n}, 1\}$, and for simplicity, assume throughout the proof that n is even (this is convenient since in this case we can refer to the 0 state. Whenever n is odd, the same proof holds by letting $\frac{1}{n}$ take the role of the 0 state).

The following notation will be useful. We define

$$\mathcal{X}[a, b] := \{x \in \mathcal{X} : a \leq x \leq b\} ,$$

and similarly define $\mathcal{X}(a, b)$, etc. accordingly. For all $x \in \mathcal{X}$, let p_x, q_x, h_x denote the transition probabilities of S_t to the right, to the left and to itself from the state x, that is:

$$p_x := P_M\left(x, x + \tfrac{2}{n}\right) = \frac{1-x}{2} \cdot \frac{1 + \tanh(\beta(x + n^{-1}))}{2} ,$$
$$q_x := P_M\left(x, x - \tfrac{2}{n}\right) = \frac{1+x}{2} \cdot \frac{1 - \tanh(\beta(x - n^{-1}))}{2} ,$$
$$h_x := P_M\left(x, x\right) = 1 - p_x - q_x .$$

By well known results on birth-and-death chains (see, e.g., [54]), the resistance r_x and conductance c_x of the edge $(x, x + 2/n)$, and the conductance c'_x of the self-loop of vertex x for $x \in \mathcal{X}[0, 1]$ are (the negative parts can be obtained immediately by symmetry)

$$r_x = \prod_{y \in \mathcal{X}(0,x]} \frac{q_y}{p_y} \ , \quad c_x = \prod_{y \in \mathcal{X}(0,x]} \frac{p_y}{q_y} \ , \quad c'_x = \frac{h_x}{p_x + q_x}(c_{x-2/n} + c_x) \ , \tag{2.6.1}$$

and the commute-time between x and y, $C_{x,y}$ for $x < y$ (the minimal time it takes the chain,

starting from x, to hit y then return to x) satisfies

$$\mathbf{E}C_{x,y} = 2c_S R(x \leftrightarrow y) \; , \tag{2.6.2}$$

where

$$c_S := \sum_{x \in \mathcal{X}} (c_x + c'_x) \quad \text{and} \quad R(x \leftrightarrow y) := \sum_{z \in \mathcal{X}[x,y)} r_z \; .$$

Our first goal is to estimate the expected commute time between 0 and ζ. This is incorporated in the next lemma.

Lemma 2.6.1. *The expected commute time between 0 and ζ has order*

$$t_{\exp} := \frac{n}{\delta} \exp\left(\frac{n}{2}\int_0^\zeta \log\frac{1+g(x)}{1-g(x)}\right) dx \; , \tag{2.6.3}$$

where $g(x) := (\tanh(\beta x) - x) / (1 - x\tanh(\beta x))$. In particular, in the special case $\delta \to 0$ we have $\mathbf{E}C_{0,\zeta} = \frac{n}{\delta}\exp\left((\frac{3}{4} + o(1))\delta^2 n\right)$, where the $o(1)$-term tends to 0 as $n \to \infty$.

Remark. If $\zeta \notin \mathcal{X}$, instead we simply choose a state in $\mathcal{X}$ which is the nearest possible to ζ. For a sufficiently large n, such a negligible adjustment would keep our calculations and arguments in tact. For the convenience of notation, let ζ denote the mentioned state in this case as well.

To prove Lemma 2.6.1, we need the following two lemmas, which establish the order of the total conductance and effective resistance respectively.

Lemma 2.6.2. *The total conductance satisfies*

$$c_S = \Theta\left(\sqrt{\frac{n}{\delta}}\exp\left(\frac{n}{2}\int_0^\zeta \log\left(\frac{1+g(x)}{1-g(x)}\right) dx\right)\right) \; .$$

Lemma 2.6.3. *The effective resistance between 0 and ζ satisfies*

$$R(0 \leftrightarrow \zeta) = \Theta(\sqrt{n/\delta}) \; .$$

Proof of Lemma 2.6.2. Notice that for any $x \in \mathcal{X}$, the holding probability h_x is uniformly bounded from below, and thus c'_x can be uniformly bounded from above by $(c_x + c_{x-2/n})$. It therefore follows that $c_S = \Theta(\tilde{c}_S)$ where $\tilde{c}_S := \sum_{x \in \mathcal{X}} c_x$, and it remains to determine $\tilde{c}_S$. We first locate the maximal edge conductance and determine its order, by means of classical

analysis.

$$\log c_x = \sum_{y \in \mathcal{X}(0,x]} \log \frac{p_y}{q_y} = \sum_{y \in \mathcal{X}(0,x]} \log \left(\frac{1-y}{1+y} \cdot \frac{1+\tanh(\beta(y+n^{-1}))}{1-\tanh(\beta(y-n^{-1}))} \right)$$

$$= \sum_{y \in \mathcal{X}(0,x]} \log \left(\frac{1+g(y)}{1-g(y)} + O(1/n) \right) = \sum_{y \in \mathcal{X}(0,x]} \log \left(\frac{1+g(y)}{1-g(y)} \right) + O(x) \qquad (2.6.4)$$

Note that $g(x)$ has a unique positive root at $x = \zeta$, and satisfies $g(x) > 0$ for $x \in (0, \zeta)$ and $g(x) < 0$ for $x > \zeta$. Therefore,

$$\log c_x \leq \log c_\zeta + O(x) \leq \log c_\zeta + O(1) ,$$

thus we move to estimate c_ζ. As $\log c_\zeta$ is simply a Riemann sum (after an appropriate rescaling), we deduce that

$$\log c_\zeta = \sum_{x \in \mathcal{X}(0,\zeta]} \log \left(\frac{1+g(x)}{1-g(x)} \right) + O(1) = \frac{n}{2} \int_0^\zeta \log \left(\frac{1+g(x)}{1-g(x)} \right) dx + O(1) ,$$

and therefore

$$c_\zeta = \Theta \left(\exp \left(\frac{n}{2} \int_0^\zeta \log \left(\frac{1+g(x)}{1-g(x)} \right) dx \right) \right) , \qquad (2.6.5)$$

$$c_x = O(c_\zeta) . \qquad (2.6.6)$$

Next, consider the ratio $c_{x+2/n}/c_x$; whenever $x \leq \zeta$, $g(x) \geq 0$, hence we have

$$\frac{c_{x+2/n}}{c_x} = \frac{p_{x+2/n}}{q_{x+2/n}} \geq \frac{1+g(x)}{1-g(x)} - O(1/n) \geq 1 + 2g(x) - O(1/n) .$$

Whenever $\frac{1}{\sqrt{\delta n}} \leq x \leq \zeta - \frac{1}{\sqrt{\delta n}}$ (using the Taylor expansions around 0 and around ζ) we obtain that $\tanh(\beta x) - x \geq \frac{1}{2}\sqrt{\delta/n}$. Combining this with the fact that $x\tanh(\beta x)$ is always non-negative, we obtain that for any such x, $2g(x) \geq \sqrt{\delta/n}$. Therefore, setting

$$\xi_1 := \sqrt{\frac{1}{\delta n}} , \quad \xi_2 := \zeta - \sqrt{\frac{1}{\delta n}} , \quad \xi_3 := \zeta + \sqrt{\frac{1}{\delta n}} , \qquad (2.6.7)$$

we get

$$\frac{c_{x+2/n}}{c_x} \geq 1 + \sqrt{\frac{\delta}{n}} - O(1/n) \quad \text{for any } x \in \mathcal{X}[\xi_1, \xi_2] \ . \tag{2.6.8}$$

Using the fact that $\delta^2 n \to \infty$, the sum of c_x-s in the above range is at most the sum of a geometric series with a quotient $1/(1 + \frac{1}{2}\sqrt{\delta/n})$ and an initial position c_ζ:

$$\sum_{x \in \mathcal{X}[\xi_1, \xi_2]} c_x \leq 3\sqrt{\frac{n}{\delta}} \cdot c_\zeta \ . \tag{2.6.9}$$

We now treat $x \geq \xi_3$; since $g(\zeta) = 0$ and $g(x)$ is decreasing for any $x \geq \zeta$, then in particular whenever $\zeta + \sqrt{\delta/n} \leq x \leq 1$ we have $-1 = g(1) \leq g(x) \leq 0$, and therefore

$$\frac{c_{x+2/n}}{c_x} = \frac{p_{x+2/n}}{q_{x+2/n}} \leq 1 + g(x) + O(1/n) \ .$$

Furthermore, for any $\zeta + \sqrt{\delta/n} \leq x \leq 1$ (using Taylor expansion around ζ) we have $\tanh(\beta x) - x \leq -\sqrt{\delta/n}$, and hence $g(x) \leq -\sqrt{\delta/n}$. We deduce that

$$\frac{c_{x+2/n}}{c_x} \leq 1 - \sqrt{\frac{\delta}{n}} + O(1/n) \quad \text{for any } x \in \mathcal{X}[\xi_3, 1] \ ,$$

and therefore

$$\sum_{x \in \mathcal{X}[\xi_3, 1]} c_x \leq 2\sqrt{\frac{n}{\delta}} \cdot c_\zeta \ . \tag{2.6.10}$$

Combining (2.6.9) and (2.6.10) together, and recalling (2.6.6), we obtain that

$$\tilde{c}_S = \sum_{x \in \mathcal{X}} c_x \leq 2 \left(|\mathcal{X}[0, \xi_1]| + 5\sqrt{n/\delta} + |\mathcal{X}[\xi_2, \xi_3]| \right) c_\zeta = O\left(\sqrt{\frac{n}{\delta}} c_\zeta \right) \ .$$

Finally, consider $x \in \mathcal{X}[\xi_2, \xi_3]$; an argument similar to the ones above (i.e., perform Taylor expansion around ζ and bound the ratio of $c_{x+2/n}/c_x$) shows that c_x is of order c_ζ in this region. This implies that for some constant $b > 0$

$$\tilde{c}_S \geq \sum_{x \in \mathcal{X}[\xi_2, \xi_2]} c_x \geq b|\mathcal{X}[\xi_2, \xi_3]|c_\zeta \geq b\sqrt{\frac{n}{\delta}} c_\zeta \ , \tag{2.6.11}$$

and altogether, plugging in (2.6.5), we get

$$\tilde{c}_S = \Theta\left(\sqrt{\frac{n}{\delta}}\exp\left(\frac{n}{2}\int_0^\zeta \log\left(\frac{1+g(x)}{1-g(x)}\right)dx\right)\right) . \tag{2.6.12}$$

∎

Proof of Lemma 2.6.3. Translating the conductances, as given in (2.6.8), to resistances, we get

$$\frac{r_{x+2/n}}{r_x} \le 1 - \sqrt{\frac{\delta}{n}} - O(1/n) \quad \text{for any } x \in \mathcal{X}[\xi_1,\xi_2] ,$$

and hence

$$\sum_{x\in\mathcal{X}[\xi_1,\xi_2]} r_x \le r_{\xi_1} 2\sqrt{n/\delta} \le 2\sqrt{n/\delta} ,$$

where in the last inequality we used the fact that $r_x \le r_{x-2/n}$ ($\le r_0 = 1$) for all $x \in \mathcal{X}[0,\zeta]$, which holds since $q_x \le p_x$ for such x. Altogether, we have the following upper bound:

$$R(0 \leftrightarrow \zeta) = \sum_{x\in\mathcal{X}[0,\xi_1]} r_x + \sum_{x\in\mathcal{X}[\xi_1,\xi_2]} r_x + \sum_{x\in\mathcal{X}[\xi_2,\zeta]} r_x$$

$$\le |\mathcal{X}[0,\xi_1]| + 2\sqrt{\frac{n}{\delta}} + |\mathcal{X}[\xi_2,\zeta]| \le 4\sqrt{n/\delta} . \tag{2.6.13}$$

For a lower bound, consider $x \in \mathcal{X}[0,\xi_1]$. Clearly, for any $x \le \frac{1}{\sqrt{\delta n}}$ we have $g(x) = \frac{\tanh(\beta x - x)}{1 - x\tanh(\beta x)} \le 2\delta x$, and hence

$$\frac{r_{x+2/n}}{r_x} = \frac{1-g(x)}{1+g(x)} + O(1/n) \ge 1 - 5x\delta \ge \exp(-6x\delta) ,$$

yielding that

$$r_{\xi_1} \ge \exp\left(-3\delta n \cdot \xi_1^2\right) \ge e^{-3} .$$

Altogether,

$$R(0 \leftrightarrow \zeta) \ge e^{-3}|\mathcal{X}[0,\xi_1]| \ge e^{-4}\sqrt{n/\delta} ,$$

and combining this with (2.6.13) we deduce that $R(0 \leftrightarrow \zeta) = \Theta(\sqrt{n/\delta})$. ∎

Proof of Lemma 2.6.1. Plugging in the estimates for $\tilde{c}_S$ and $R(0 \leftrightarrow \zeta)$ in (2.6.2), we get

$$\mathbf{E}C_{0,\zeta} = \Theta\left(\frac{n}{\delta}\exp\left(\frac{n}{2}\int_0^\zeta \log\left(\frac{1+g(x)}{1-g(x)}\right)dx\right)\right) . \tag{2.6.14}$$

This completes the proof of the lemma 2.6.1. ∎

Note that by symmetry, the expected hitting time from ζ to $-\zeta$ is exactly the expected commute time between 0 and ζ. Hence,

$$\mathbf{E}_\zeta[\tau_{-\zeta}] = \Theta(t_{\exp}) . \tag{2.6.15}$$

In order to show that the above hitting time is the leading order term in the mixing-time at low temperatures, we need the following lemma, which addresses the order of the hitting time from 1 to ζ.

Lemma 2.6.4. *The normalized magnetization chain S_t in the low temperature regimes satisfies $\mathbf{E}_1\tau_\zeta = o(t_{\exp})$.*

Proof. First consider the case where δ is bounded below by some constant. Notice that, as $p_x \leq q_x$ for all $x \geq \zeta$, in this region S_t is a supermartingale. Therefore, Lemma 2.3.5 (or simply standard results on the simple random walk, which dominates our chain in this case) implies that $\mathbf{E}_1\tau_\zeta = O(n^2)$. Combining this with the fact that $t_{\exp} \geq \exp(cn)$ for some constant c in this case, we immediately obtain that $\mathbf{E}_1\tau_\zeta = o(t_{\exp})$.

Next, assume that $\delta = o(1)$. Note that in this case, the Taylor expansion $\tanh(\beta x) = \beta x - \frac{1}{3}(\beta x)^3 + O((\beta x)^5)$ implies that

$$\zeta = \sqrt{3\delta/\beta^3 - O((\beta\zeta))^5} = \sqrt{3\delta} + O(\delta^{3/2}) . \tag{2.6.16}$$

Recalling that $\mathbf{E}[S_{t+1} \mid S_t = s] \leq s + \frac{1}{n}(\tanh(\beta s) - s)$ (as $s \geq 0$), Jensen's inequality (using the concavity of the Hyperbolic tangent) gives

$$\mathbf{E}[S_{t+1} - S_t] = \mathbf{E}(\mathbf{E}[S_{t+1} - S_t \mid S_t]) \leq \frac{1}{n}\left(\mathbf{E}\tanh(\beta S_t) - \mathbf{E}S_t\right)$$

$$\leq \frac{1}{n}\left(\tanh(\beta\mathbf{E}S_t) - \mathbf{E}S_t\right) . \tag{2.6.17}$$

Further note that the function $\tanh(\beta s)$ has the following Taylor expansion around ζ (for some ξ between s and ζ):

$$\tanh(\beta s) = \zeta + \beta(1 - \zeta^2)(s - \zeta) + \beta^2(-1 + \zeta^2)\zeta(s - \zeta)^2$$

$$+ \frac{\beta^3}{3}(-1 + 4\zeta^2 - \zeta^4)(s - \zeta)^3 + \frac{\tanh^{(4)}(\xi)}{4!}(s - \zeta)^4 . \tag{2.6.18}$$

Since $\tanh^{(4)}(x) < 5$ for any $x \geq 0$, (2.6.18) implies that for a sufficiently large n the term $-\frac{1}{3}(s-\zeta)^3$ absorbs the last term in the expansion (2.6.18). Together with (2.6.16), we obtain

that

$$\tanh(\beta s) \leq \zeta + \beta(1 - \zeta^2)(s - \zeta) + \beta^2(-1 + \zeta^2)\sqrt{\delta}(s - \zeta)^2 \ .$$

Therefore, (2.6.17) follows:

$$\mathbf{E}[S_{t+1} - S_t] \leq -\frac{\sqrt{\delta}}{2n}(\mathbf{E}S_t - \zeta)^2 \ . \tag{2.6.19}$$

Set

$$b_i = 2^{-i} \ , i_2 = \min\{i : b_i < \sqrt{\delta}\} \text{ and } u_i = \min\{t : \mathbf{E}S_t - \zeta < b_i\} \ ,$$

noting that this gives $b_i/2 \leq \mathbf{E}S_t - \zeta \leq b_i$ for any $t \in [u_i, u_{i+1}]$. It follows that

$$u_{i+1} - u_i \leq \frac{b_i/2}{\frac{\sqrt{\delta}}{2n}(\frac{b_i}{2})^2} = \frac{4n}{\sqrt{\delta}b_i} \ ,$$

and hence

$$\sum_{i=1}^{i_2} u_{i+1} - u_i \leq \sum_{i:b_i^2 > \delta} \frac{4n}{\sqrt{\delta}b_i} = O(n/\delta) \ ,$$

where we used the fact that the series $\{b_i^{-1}\}$ is geometric with ratio 2. We claim that this implies the required bound on $\mathbf{E}_1\tau_\zeta$. To see this, recall (2.6.19), according to which $W_t := n(S_t - \zeta)\mathbf{1}_{\{\tau_\zeta > t\}}$ is a supermartingale with bounded increments, whose variance is uniformly bounded from below on the event $\tau_\zeta > t$ (as the holding probabilities of (S_t) are uniformly bounded from above, see (2.3.4)). Moreover, the above argument gives $\mathbf{E}W_t \leq n\sqrt{\delta}$ for some $t = O(n/\delta)$. Thus, applying Lemma 2.3.5 and taking expectation, we deduce that $\mathbf{E}_1\tau_\zeta = O(n/\delta + \delta n^2) = O(\delta n^2)$, which in turns gives $\mathbf{E}_1\tau_\zeta = o(t_{\exp})$. ∎

Remark. With additional effort, we can establish that $\mathbf{E}_0\tau_{\pm\zeta} = o(t_{\exp})$ (for more details, see [26]), where $\tau_{\pm\zeta} = \min\{t : |S_t| \geq \zeta\}$. By combining this with the of S_t symmetry and applying the geometric trial method, we can obtain the expected commute time between 0 and ζ:

$$\mathbf{E}_\zeta\tau_0 = (\tfrac{1}{2} + o(1))\mathbf{E}C_{0,\zeta} = \Theta(t_{\exp}) \ ,$$

and therefore conclude that $\mathbf{E}_1\tau_0 = \Theta(t_{\exp})$.

Upper bound for mixing

Combining Lemma 2.6.4 and (2.6.15), we conclude that $\mathbf{E}_1 \tau_{-\zeta} = \Theta(t_{\exp})$ and hence $\mathbf{E}_1 \tau_0 = O(t_{\exp})$. Together with Lemma 2.3.1, this implies that the magnetization chain will coalescence in $O(t_{\exp})$ steps with probability arbitrarily close to 1. At this point, Lemma 2.3.3 immediately gives that the Glauber dynamics achieves full mixing within $O(n \log n)$ additional steps. The following simple lemma thus completes the proof of the upper bound for the mixing time.

Lemma 2.6.5. *Let $t_{\exp}$ be as defined in Lemma 2.6.1. Then $n \log n = o(t_{\exp})$.*

Proof. In case $\delta \geq c > 0$ for some constant c, we have $t_{\exp} \geq n \exp(c'n)$ for some constant $c' > 0$ and hence $n \log n = o(t_{\exp})$. It remains to treat the case $\delta = o(1)$.

Suppose first that $\delta = o(1)$ and $\delta \geq cn^{-1/3}$ for some constant $c > 0$. In this case, we have $t_{\exp} = \frac{n}{\delta} \exp\left(\left(\frac{3}{4} + o(1) \right) \delta^2 n \right)$ and thus $n \exp(\frac{1}{2} n^{1/3}) = O(t_{\exp})$, giving $n \log n = o(t_{\exp})$. Finally, if $\delta = o(n^{-1/3})$, we can simply conclude that $n^{4/3} = O(t_{\exp})$ and hence $n \log n = o(t_{\exp})$. ∎

Lower bound for mixing

The lower bound will follow from showing that the probability of hitting $-\zeta$ within $\varepsilon t_{\exp}$ steps is small, for some small $\varepsilon > 0$ to be chosen later. To this end, we need the following simple lemma:

Lemma 2.6.6. *Let X denote a Markov chain over some finite state space Ω, $y \in \Omega$ denote a target state, and T be an integer. Further let $x \in \Omega$ denote the state with the smallest probability of hitting y after at most T steps, i.e., x minimizes $\mathbf{P}_x(\tau_y \leq T)$. The following holds:*

$$\mathbf{P}_x(\tau_y \leq T) \leq \frac{T}{\mathbf{E}_x \tau_y} \ .$$

Proof. Set $p = \mathbf{P}_x(\tau_y \leq T)$. By definition, $\mathbf{P}_z(\tau_y \leq T) \geq p$ for all $z \in \Omega$, hence the hitting time from x to y is stochastically dominated by a geometric random variable with success probability p, multiplied by T. That is, we have $\mathbf{E}_x \tau_y \leq T/p$, completing the proof. ∎

The final fact we would require is that the stationary probability of $\mathcal{X}[-1, -\zeta]$ is strictly positive. This is stated by the following lemma.

Lemma 2.6.7. *There exists some absolute constant $0 < C_\pi < 1$ such that*

$$C_\pi \leq \pi(\mathcal{X}[\zeta, 1]) \ (= \pi(\mathcal{X}[-1, -\zeta])) \ .$$

Proof. Repeating the derivation of (2.6.11), we can easily get

$$c_{\mathcal{X}[\zeta,1]} := \sum_{x \in \mathcal{X}[\zeta,1]} (c_x + c'_x) \geq \Theta\left(\sqrt{\frac{n}{\delta}} \exp\left(\frac{n}{2} \int_0^\zeta \log \frac{1+g(x)}{1-g(x)} \right) dx \right) \ .$$

Combining the above bound with (2.6.12), we conclude that there exists some $C_\pi > 0$, such that $\pi(\mathcal{X}[\zeta,1]) \geq C_\pi$. ∎

Plugging in the target state $-\zeta$ into Lemma 2.6.6, and recalling that the monotone-coupling implies that, for any T, the initial state $s_0 = 1$ has the smallest probability (among all initial states) of hitting $-\zeta$ within T steps, we deduce that, for a sufficiently small $\varepsilon > 0$,

$$\mathbf{P}_1(\tau_{-\zeta} \leq \varepsilon t_{\exp}) \leq \frac{1}{2} C_\pi \ .$$

This implies that

$$t_{\exp} = O\left(t_{\text{MIX}}\left(\frac{1}{2} C_\pi \right) \right) ,$$

which in turn gives

$$t_{\exp} = O\left(t_{\text{MIX}}\left(\tfrac{1}{4} \right) \right) \ .$$

2.6.2 Spectral gap analysis

The lower bound is straightforward (as the relaxation time is always at most the mixing time) and we turn to prove the upper bound. Note that, by Lemma 2.6.7, we have $\pi(\mathcal{X}[\zeta,1]) \geq C_\pi > 0$. Suppose first that $\texttt{gap} \cdot t_{\text{MIX}}(\frac{1}{4}) \to \infty$. In this case, one can apply Lemma 2.3.7 onto the birth-and-death chain (S_t), with a choice of $\alpha = \pi(\mathcal{X}[\zeta,1])$ and $\beta = 1 - \pi(\mathcal{X}[\zeta,1])$ (recall that $t_{\text{MIX}}(\frac{1}{4}) = \Theta(\mathbf{E}_1 \tau_{-\zeta})$). It follows that

$$\mathbf{E}_\zeta \tau_{-\zeta} = o\left(\mathbf{E}_1 \tau_{-\zeta} \right) \ .$$

However, as both quantities above should have the same order as $t_{\text{MIX}}(\frac{1}{4})$, this leads to a contradiction. We therefore have $\texttt{gap} \cdot t_{\text{MIX}}(\frac{1}{4}) = O(1)$, completing the proof of the upper bound.

Chapter 3

Mixing for the Ising-model on regular trees at criticality

3.1 Introduction

In the classical Ising model, the underlying geometry is the d-dimensional lattice, and there is a critical inverse-temperature β_c where the static Gibbs measure exhibits a phase transition with respect to long-range correlations between spins. While the main focus of the physics community is on critical behavior (see the 20 volumes of [30]), so far, most of the rigorous mathematical analysis was confined to the non-critical regimes.

Supported by many experiments and studies in the theory of dynamical critical phenomena, physicists believe that the spectral-gap of the continuous-time dynamics on lattices has the following *critical slowing down* behavior (e.g., [44, 50, 64, 84]): At high temperatures $(\beta < \beta_c)$ the inverse-gap is $O(1)$, at the critical β_c it is polynomial in the surface-area and at low temperatures it is exponential in it. This is known for $\mathbb{Z}^2$ except at the critical β_c, and establishing the order of the gap at criticality seems extremely challenging. In fact, the only underlying geometry, where the critical behavior of the spectral-gap has been fully established, is the complete graph (see [25]).

The important case of the Ising model on a regular tree, known as the Bethe lattice, has been intensively studied (e.g., [8, 9, 13–15, 36, 45, 46, 62, 66, 73]). We recall the definition for the Ising model as in (1.1.1), and we assume $J \equiv 1$ throughout the chapter. On this canonical example of a *non-amenable* graph (one whose boundary is proportional to its volume), the model exhibits a rich behavior. For example, it has two distinct critical inverse-temperatures: one for uniqueness of the Gibbs state, and another for the purity of the free-boundary state. The latter, β_c, coincides with the critical spin-glass parameter.

As we later describe, previous results on the Ising model on a regular tree imply that the correct parameter to play the role of the surface-area is the tree-height h: It was shown that the inverse-gap is $O(1)$ for $\beta < \beta_c$ and exponential in h for $\beta > \beta_c$, yet its critical behavior remained unknown.

In this chapter, we complete the picture of the spectral-gap of the dynamics for the critical Ising model on a regular tree, by establishing that it is indeed polynomial in h. Furthermore, this holds under any boundary condition, and an analogous result is obtained for the L^1 (total-variation) mixing time, denoted by t_{MIX} (formally defined in Subsection 3.2.1).

Theorem 6. *Fix $b \geq 2$ and let $\beta_c = \mathrm{arctanh}(1/\sqrt{b})$ denote the critical inverse-temperature for the Ising model on the b-ary tree of height h. Then there exists some constant $c > 0$ independent of b, so that the following holds: For any boundary condition τ, the continuous-time Glauber dynamics for the above critical Ising model satisfies $\mathtt{gap}^{-1} \leq t_{\mathrm{MIX}} = O(h^c)$.*

One of the main obstacles in proving the above result is the arbitrary boundary condition, due to which the spin system loses its symmetry (and the task of analyzing the dynamics becomes considerably more involved). Note that, although boundary conditions are believed to only accelerate the mixing of the dynamics, even tracking the effect of the (symmetric) all-plus boundary on lattices for $\beta > \beta_c$ is a formidable open problem (see [65]).

In light of the above theorem and the known fact that the inverse-gap is exponential in h at low temperatures ($\beta > \beta_c$ fixed), it is natural to ask how the transition between these two phases occurs, and in particular, what the critical exponent of $\beta - \beta_c$ is. This is answered by the following theorem, which establishes that $\log(\mathtt{gap}^{-1}) \asymp (\beta - \beta_c)h + \log h$ for small $\beta - \beta_c$. Moreover, this result also holds for $\beta = \beta_c + o(1)$, and thus pinpoints the transition to a polynomial inverse-gap at $\beta - \beta_c \asymp \frac{\log h}{h}$.

Theorem 7. *For some $\varepsilon_0 > 0$, any $b \geq 2$ fixed and all $\beta_c < \beta < \beta_c + \varepsilon_0$, where $\beta_c = \mathrm{arctanh}(1/\sqrt{b})$ is the critical spin-glass parameter, the following holds: The continuous-time Glauber dynamics for the Ising model on a b-ary tree with inverse-temperature β and free boundary satisfies*

$$\begin{aligned}
\mathtt{gap}^{-1} &= h^{\Theta(1)} & \text{if } \beta = \beta_c + O(\tfrac{\log h}{h}) \ , \\
\mathtt{gap}^{-1} &= \exp\left[\Theta\left((\beta - \beta_c)h\right)\right] & \text{otherwise.}
\end{aligned} \tag{3.1.1}$$

Furthermore, both upper bounds hold under any boundary condition τ, and (3.1.1) remains valid if $\mathtt{gap}^{-1}$ is replaced by t_{MIX}.

In the above theorem and in what follows, the notation $f = \Theta(g)$ stands for $f = O(g)$ and $g = O(f)$.

Finally, our results include new lower bounds on the critical inverse-gap and the total-variation mixing-time (see Theorem 8). The lower bound on $\texttt{gap}^{-1}$ refutes a conjecture of [8], according to which the continuous-time inverse-gap is linear in h. Our lower bound on t_{MIX} is of independent interest: Although in our setting the ratio between t_{MIX} and $\texttt{gap}^{-1}$ is at most poly-logarithmic in n, the number of sites, we were able to provide a lower bound of order $\log n$ on this ratio without resorting to eigenfunction analysis.

3.1.1   Background

The thoroughly studied question of whether the free boundary state is *pure* (or *extremal*) in the Ising model on the Bethe lattice can be formulated as follows: Does the effect that a typical boundary has on the spin at the root vanish as the size of the tree tends to infinity? It is well-known that one can sample a configuration for the tree according to the Gibbs distribution with free boundary by propagating spins along the tree (from a site to its children) with an appropriate bias (see Subsection 3.2.2 for details). Hence, the above question is equivalent to asking wether the spin at the root can be reconstructed from its leaves, and as such has applications in Information Theory and Phylogeny (see [36] for further details).

In sufficiently high temperatures, there is a unique Gibbs state for the Ising model on a b-ary tree ($b \geq 2$), hence in particular the free boundary state is pure. The phase-transition with respect to the uniqueness of the Gibbs distribution occurs at the inverse-temperature $\beta_u = \text{arctanh}(1/b)$, as established in 1974 by Preston [77].

In [14], the authors studied the critical spin-glass model on the Bethe lattice (see also [13, 15]), i.e., the Ising model with a boundary of i.i.d. uniform spins. Following that work, it was finally shown in [9] that the phase-transition in the free-boundary extremality has the same critical inverse-temperature as in the spin-glass model, $\beta_c = \text{arctanh}(1/\sqrt{b})$. That is, the free-boundary state is pure iff $\beta \leq \beta_c$. This was later reproved in [45, 46].

The inverse-gap of the Glauber dynamics for the Ising model on a graph G was related in [8] to the *cut-width* of the graph, $\xi(G)$, defined as follows: It is the minimum integer m, such that for some labeling of the vertices $\{v_1, \ldots, v_n\}$ and any $k \in [n]$, there are at most m edges between $\{v_1, \ldots, v_k\}$ and $\{v_{k+1}, \ldots, v_n\}$. The authors of [8] proved that for any bounded degree graph G, the continuous-time gap satisfies $\texttt{gap}^{-1} = \exp[O(\xi(G)\beta)]$.

Recalling the aforementioned picture of the phase-transition of the gap, this supports the claim that the cut-width is the correct extension of the surface-area to general graphs. One can easily verify that for $\mathbb{Z}_L^d$ (the d-dimensional box of side-length L) the cut-width has the same order as the surface-area L^{d-1}, while for a regular tree of height h it is of order h.

Indeed, for the Ising model on a b-ary tree with h levels and free boundary, it was shown

in [8] that the inverse-gap is $O(1)$ for all $\beta < \beta_c$, whereas for $\beta > \beta_c$ it satisfies $\log \mathtt{gap}^{-1} \asymp h$ (with constants that depend on b and β). The behavior of the gap at criticality was left as an open problem: it is proved in [8] that the critical $\mathtt{gap}^{-1}$ is at least linear in h and conjectured that this is tight. A weaker conjecture of [8] states that $\mathtt{gap}^{-1} = \exp(o(h))$.

Further results on the dynamics were obtained in [66], showing that the log-Sobolev constant $\alpha_{\mathbf{s}}$ (defined in Section 5.1) is uniformly bounded away from zero for $\beta < \beta_c$ in the free-boundary case, as well as for any β under the all-plus boundary condition. While this implies that $\mathtt{gap}^{-1} = O(1)$ in these regimes, it sheds no new light on the behavior of the parameters $\mathtt{gap}, \alpha_{\mathbf{s}}$ in our setting of the critical Ising model on a regular tree with free-boundary.

3.1.2 The critical inverse-gap and mixing-time

Theorems 6,7, stated above, establish that on a regular tree of height h, the critical and near-critical continuous-time $\mathtt{gap}^{-1}$ and t_{MIX} are polynomial in h. In particular, this confirms the conjecture of [8] that the critical inverse-gap is $\exp(o(h))$.

Moreover, our upper bounds hold for any boundary condition, while matching the behavior of the free-boundary case: Indeed, in this case the critical inverse-gap is polynomial in h (as [8] showed it is at least linear), and for $\beta - \beta_c > 0$ small we do have that $\log(\mathtt{gap}^{-1}) \asymp (\beta - \beta_c)h$. For comparison, recall that under the all-plus boundary condition, [66] showed that $\mathtt{gap}^{-1} = O(1)$ at all temperatures.

We next address the conjecture of [8] that the critical inverse-gap is in fact linear in h. The proof that the critical $\mathtt{gap}^{-1}$ has order at least h uses the same argument that gives a tight lower bound at high temperatures: Applying the Dirichlet form (see Subsection 3.2.4) to the sum of spins at the leaves as a test-function. Hence, the idea behind the above conjecture is that the sum of spins at the boundary (that can be thought of as the magnetization) approximates the second eigenfunction also for $\beta = \beta_c$.

The following theorem refutes this conjecture, and en-route also implies that $\alpha_{\mathbf{s}} = o(1)$ at criticality. In addition, this theorem provides a nontrivial lower bound on t_{MIX} that separates it from $\mathtt{gap}^{-1}$ (thus far, our bounds in Theorems 6,7 applied to both parameters as one).

Theorem 8. *Fix $b \geq 2$ and let $\beta_c = \mathrm{arctanh}(1/\sqrt{b})$ be the critical inverse-temperature for the Ising model on the b-ary tree with n vertices. Then the corresponding discrete-time Glauber dynamics with free boundary satisfies:*

$$\mathtt{gap}^{-1} \geq c_1 \, n \, (\log n)^2 \; , \tag{3.1.2}$$

$$t_{\mathrm{MIX}} \geq c_2 \, n \, (\log n)^3 \; , \tag{3.1.3}$$

for some $c_1, c_2 > 0$. Furthermore, $t_{\text{MIX}} \geq c \, \texttt{gap}^{-1} \log n$ for some $c > 0$.

Indeed, the above theorem implies that in continuous-time, $\texttt{gap}^{-1}$ has order at least h^2 and t_{MIX} has order at-least h^3, where h is again the height of the tree. By known facts on the log-Sobolev constant (see Section 5.1, Corollary 3.2.4), in our setting we have $t_{\text{MIX}} = O(\alpha_{\mathbf{s}}^{-1} h)$, and it then follows that $\alpha_{\mathbf{s}} = O(h^{-2}) = o(1)$.

We note that by related results on the log-Sobolev constant, it follows that in the Ising model on a regular tree, for any temperature and with any boundary condition we have $t_{\text{MIX}} = O(\texttt{gap}^{-1} \log^2 n)$. In light of this, establishing a lower bound of order $\log n$ on the ratio between t_{MIX} and $\texttt{gap}^{-1}$ is quite delicate (e.g., proving such a bound usually involves constructing a distinguishing statistic via a suitable eigenfunction (Wilson's method [85])).

3.1.3   Techniques and proof ideas

To prove the main theorem, our general approach is a recursive analysis of the spectral-gap via an appropriate block-dynamics (roughly put, multiple sites comprising a block are updated simultaneously in each step of this dynamics; see Subsection 3.2.5 for a formal definition). This provides an estimate of the spectral-gap of the single-site dynamics in terms of those of the individual blocks and the block-dynamics chain itself (see [64]). However, as opposed to most applications of the block-dynamics method, where the blocks are of relatively small size, in our setting we must partition a tree of height h to subtrees of height linear in h. This imposes arbitrary boundary conditions on the individual blocks, and highly complicates the analysis of the block-dynamics chain.

In order to estimate the gap of the block-dynamics chain, we apply the method of Decomposition of Markov chains, introduced in [48] (for details on this method see Subsection 3.2.6). Combining this method with a few other ideas (such as establishing contraction and controlling the external field in certain chains), the proof of Theorem 6 is reduced into the following spatial-mixing/reconstruction type problem. Consider the procedure, where we assign the spins of the boundary given the value at the root of the tree, then reconstruct the root from the values at the boundary. The key quantity required in our setting is the difference in the expected outcome of the root, comparing the cases where its initial spin was either positive or negative.

This quantity was studied by [73] in the free-boundary case, where it was related to capacity-type parameters of the tree (see [36] for a related result corresponding to the high temperature regime). Unfortunately, in our case we have an arbitrary boundary condition, imposed by the block-dynamics. This eliminates the symmetry of the system, which was a crucial part of the arguments of [73]. The most delicate step in the proof of Theorem 6 is the extension of these results of [73] to any boundary condition. This is achieved by

carefully tracking down the effect of the boundary on the expected reconstruction result in each site, combined with correlation inequalities and an analytical study of the corresponding log-likelihood-ratio function.

The lower bound on the critical inverse-gap reflects the change in the structure of the dominant eigenfunctions between high and low temperatures. At high temperatures, the sum of spins on the boundary gives the correct order of the gap. At low temperatures, a useful lower bound on $\mathtt{gap}^{-1}$ was shown in [8] via the recursive-majority function (intuitively, this reflects the behavior at the root: Although this spin may occasionally flip its value, at low temperature it constantly tends to revert to its biased state). Our results show that at criticality, a lower bound improving upon that of [8] is obtained by essentially merging the above two functions into a weighted sum of spins, where the weight of a spin is determined by its tree level.

To establish a lower bound on t_{MIX} of order $\mathtt{gap}^{-1}h$, we consider a certain *speed-up* version of the dynamics: a block-dynamics, whose blocks are a mixture of singletons and large subtrees. The key ingredient here is the Censoring Inequality of Peres and Winkler [75], that shows that this dynamics indeed mixes as fast as the usual (single-site) one. We then consider a series of modified versions of this dynamics, and study their mixing with respect to the total-variation and Hellinger distances. In the end, we arrive at a product chain, whose components are each the single-site dynamics on a subtree of height linear in h. This latter chain provides the required lower bound on t_{MIX}.

3.1.4   Organization

The rest of this chapter is organized as follows. Section 5.1 contains several preliminary facts and definitions. In Section 3.3 we prove a spatial-mixing type result on the critical and near-critical Ising model on a tree with an arbitrary boundary condition. This then serves as one of the key ingredients in the proof of the main result, Theorem 6, which appears in Section 3.4. In Section 3.5 we prove Theorem 8, providing the lower bounds for the critical inverse-gap and mixing-time. Section 3.6 contains the proof of Theorem 7, addressing the near-critical behavior of $\mathtt{gap}^{-1}$ and t_{MIX}. The final section, Section 3.7, is devoted to concluding remarks and open problems.

3.2 Preliminaries

3.2.1 Total-variation mixing

Let (X_t) be an aperiodic irreducible Markov chain on a finite state space Ω, with stationary distribution π. For any two distributions ϕ, ψ on Ω, the *total-variation distance* of ϕ and ψ is defined as

$$\|\phi - \psi\|_{\mathrm{TV}} := \sup_{A \subset \Omega} |\phi(A) - \psi(A)| = \frac{1}{2} \sum_{x \in \Omega} |\phi(x) - \psi(x)| \ .$$

The (worst-case) total-variation *mixing-time* of (X_t), denoted by $t_{\mathrm{MIX}}(\varepsilon)$ for $0 < \varepsilon < 1$, is defined to be

$$t_{\mathrm{MIX}}(\varepsilon) := \min \left\{ t : \max_{x \in \Omega} \|\mathbf{P}_x(X_t \in \cdot) - \pi\|_{\mathrm{TV}} \leq \varepsilon \right\} ,$$

where $\mathbf{P}_x$ denotes the probability given that $X_0 = x$. As it is easy and well known (cf., e.g., [3, Chapter 4]) that the spectral-gap of (X_t) satisfies $\mathtt{gap}^{-1} \leq t_{\mathrm{MIX}}(1/e)$, it will be convenient to use the abbreviation

$$t_{\mathrm{MIX}} := t_{\mathrm{MIX}}(1/e) \ .$$

Analogously, for a continuous-time chain on Ω with heat-kernel H_t, we define t_{MIX} as the minimum t such that $\max_{x \in \Omega} \|H_t(x, \cdot) - \pi\|_{\mathrm{TV}} \leq 1/e$.

3.2.2 The Ising model on trees

When the underlying geometry of the Ising model is a tree with free boundary condition, the Gibbs measure has a natural constructive representation. This appears in the following well known claim (see, e.g., [36] for more details).

Claim 3.2.1. *Consider the Ising model on a tree T rooted at ρ with free boundary condition and at the inverse-temperature β. For all $e \in E(T)$, let $\eta_e \in \{\pm 1\}$ be i.i.d. random variables with $\mathbf{P}(\eta_e = 1) = (1 + \tanh \beta)/2$. Furthermore, let $\sigma(\rho)$ be a uniform spin, independent of $\{\eta_e\}$, and for $v \neq \rho$,*

$$\sigma(v) = \sigma(\rho) \prod_{e \in \mathcal{P}(\rho, v)} \eta_e \ , \quad \text{where } \mathcal{P}(\rho, v) \text{ is the simple path from } \rho \text{ to } v.$$

Then the distribution of the resulting σ is the corresponding Gibbs measure.

In light of the above claim, one is able to sample a configuration according to Gibbs distribution on a tree with free boundary condition using the following simple scheme: Assign a uniform spin at the root ρ, then scan the tree from top to bottom, successively assigning

each site with a spin according to the value at its parent. More precisely, a vertex is assigned the same spin as its parent with probability $(1 + \tanh \beta)/2$, and the opposite one otherwise. Equivalently, a vertex inherits the spin of its parent with probability $\tanh \beta$, and otherwise it receives an independent uniform spin. Finally, for the conditional Gibbs distribution given a plus spin at the root ρ, we assign ρ a plus spin rather than a uniform spin, and carry on as above.

However, notice that the above *does not* hold for the Ising model in the presence of a boundary condition, which may impose a different external influence on different sites.

3.2.3 L^2-capacity

The authors of [73] studied certain spatial mixing properties of the Ising model on trees (with free or all-plus boundary conditions), and related them to the L^p-capacity of the underlying tree. In Section 3.3, we extend some of the results of [73] to the (highly asymmetric) case of a tree with an arbitrary boundary condition, and relate a certain "decay of correlation" property to the L^2-capacity of the tree, defined as follows.

Let T be a tree rooted at ρ, denote its leaves by ∂T, and throughout the chapter, write $(u, v) \in E(T)$ for the directed edge between a vertex u and its child v. We further define $\mathrm{dist}(u, v)$ as the length (in edges) of the simple path connecting u and v in T.

For each $e \in E(T)$, assign the *resistance* $R_e \geq 0$ to the edge e. We say that a non-negative function $f : E(T) \to \mathbb{R}$ is a *flow* on T if the following holds for all $(u, v) \in E(T)$ with $v \notin \partial T$:

$$f(u, v) = \sum_{(v,w) \in E(T)} f(v, w) \ ,$$

that is, the incoming flow equals the outgoing flow on each internal vertex v in T. For any flow f, define its *strength* $|f|$ and *voltage* $V(f)$ by

$$|f| := \sum_{(\rho,v) \in E(T)} f(\rho, v) \ , \ V(f) := \sup \left\{ \sum_{e \in \mathcal{P}(\rho,w)} f(e) R_e : w \in \partial T \right\} \ ,$$

where $\mathcal{P}(\rho, w)$ denotes the simple path from ρ to w. Given these definitions, we now define the L^2-capacity $\mathrm{cap}_2(T)$ to be

$$\mathrm{cap}_2(T) := \sup\{|f| : f \text{ is a flow with } V(f) \leq 1\} \ .$$

For results on the L^2-capacity of general networks (and more generally, L^p-capacities, where the expression $f(e)R_e$ in the above definition of $V(f)$ is replaced by its $(p-1)$ power), as part of Discrete Nonlinear Potential Theory, cf., e.g., [68], [81], [82] and the references therein.

For our proofs, we will use the well-known fact that the L^2-capacity of the tree T is precisely the *effective conductance* between the root ρ and the leaves ∂T, denoted by $C_{\text{eff}}(\rho \leftrightarrow \partial T)$. See, e.g., [63] for further information on electrical networks.

3.2.4 Spectral gap and log-Sobolev constant

Our bound on the mixing time of Glauber dynamics for the Ising model on trees will be derived from a recursive analysis of the spectral gap of this chain. This analysis uses spatial-mixing type results (and their relation to the above mentioned L^2 capacity) as a building block. We next describe how the mixing-time can be bounded via the spectral-gap in our setting.

The spectral gap and log-Sobolev constant of a reversible Markov chain with stationary distribution π are given by the following Dirichlet forms (see, e.g., [3, Chapter 3,8]):

$$\mathtt{gap} = \inf_f \frac{\mathcal{E}(f)}{\text{Var}_\pi(f)} \ , \qquad \alpha_{\mathbf{s}} = \inf_f \frac{\mathcal{E}(f)}{\text{Ent}(f)} \ , \tag{3.2.1}$$

where

$$\mathcal{E}(f) = \langle (I - P)f, f \rangle_\pi = \frac{1}{2} \sum_{x,y \in \Omega} [f(x) - f(y)]^2 \, \pi(x) P(x,y) \ , \tag{3.2.2}$$

$$\text{Ent}_\pi(f) = \mathbf{E}_\pi \left(f^2 \log(f^2/\mathbf{E}_\pi f^2) \right) \ . \tag{3.2.3}$$

By bounding the log-Sobolev constant, one may obtain remarkably sharp upper bounds on the L^2 mixing-time: cf., e.g., [18–21, 79]. The following result of Diaconis and Saloff-Coste [20, Theorem 3.7] (see also [79, Corollary 2.2.7]) demonstrates this powerful method; its next formulation for discrete-time appears in [3, Chapter 8]. As we are interested in total-variation mixing, we write this bound in terms of t_{MIX}, though it in fact holds also for the (larger) L^2 mixing-time.

Theorem 3.2.2 ([20], [79], reformulated)**.** *For any reversible finite Markov chain with stationary distribution π,*

$$t_{\text{MIX}}(1/\mathrm{e}) \leq \frac{1}{4}\alpha_{\mathbf{s}}^{-1} \left(\log\log(1/\pi^*) + 4 \right) \ ,$$

where $\pi^* = \min_x \pi(x)$.

We can then apply a result of [66], which provides a useful bound on $\alpha_{\mathbf{s}}$ in terms of $\mathtt{gap}$ in our setting, and obtain an upper bound on the mixing-time.

Theorem 3.2.3 ([66, Theorem 5.7]). *There exists some $c > 0$ such that the Ising model on the b-ary tree with n vertices satisfies $\alpha_{\mathtt{s}} \geq c \cdot \mathtt{gap}/\log n$.*

Note that the proof of the last theorem holds for any β and under any boundary condition. Combining Theorems 3.2.2 and 3.2.3, and noticing that $\pi_* \geq 2^{-n} \exp(-2\beta n)$ (as there are 2^n configurations, and the ratio between the maximum and minimum probability of a configuration is at most $\exp(2\beta n)$), we obtain the following corollary:

Corollary 3.2.4. *The Glauber dynamics for the Ising model on a b-ary tree with n vertices satisfies $t_{\mathrm{MIX}} = O\left(\alpha_{\mathtt{s}}^{-1} \log n\right) = O\left(\mathtt{gap}^{-1} \log^2 n\right)$ for any β and any boundary condition.*

The above corollary reduces the task of obtaining an upper bound for the mixing-time into establishing a suitable lower bound on the spectral gap. This will be achieved using a block dynamics analysis.

3.2.5 From single site dynamics to block dynamics

Consider a cover of V by a collection of subsets $\{B_1, \ldots, B_k\}$, which we will refer to as "blocks". The *block dynamics* corresponding to $B_1, \ldots, B_k$ is the Markov chain, where at each step a uniformly chosen block is updated according to the stationary distribution given the rest of the system. That is, the entire set of spins of the chosen block is updated simultaneously, whereas all other spins remain unchanged. One can verify that the block dynamics is reversible with respect to the Gibbs distribution μ_n.

Recall that, given a subset of the sites $U \subset V$, a boundary condition η imposed on U is the restriction of the sites $U^c = V \setminus U$ to all agree with η throughout the dynamics, i.e., only sites in U are considered for updates. It will sometimes be useful to consider $\eta \in \Omega$ (rather than a configuration of the sites U^c), in which case only its restriction to U^c is accounted for.

The following theorem shows the useful connection between the single-site dynamics and the block dynamics. This theorem appears in [64] in a more general setting, and following is its reformulation for the special case of Glauber dynamics for the Ising model on a finite graph with an arbitrary boundary condition. Though the original theorem is stated for the continuous-time dynamics, its proof naturally extends to the discrete-time case; we provide its details for completeness.

Proposition 3.2.5 ([64, Proposition 3.4], restated). *Consider the discrete time Glauber dynamics on a b-ary tree with boundary condition η. Let $\mathtt{gap}_U^\eta$ be the spectral-gap of the single-site dynamics on a subset $U \subset V$ of the sites, and $\mathtt{gap}_{\mathcal{B}}^\eta$ be the spectral-gap of the*

block dynamics corresponding to $B_1, \ldots, B_k$, an arbitrary cover of a vertex set $W \subset V$. The following holds:

$$\mathrm{gap}_W^\eta \geq \frac{k}{|W|} \mathrm{gap}_{\mathcal{B}}^\eta \inf_i \inf_\varphi |B_i| \mathrm{gap}_{B_i}^\varphi \left(\sup_{x \in W} \#\{i : B_i \ni x\} \right)^{-1} .$$

Proof. Let P denote the transition kernel of the above Glauber dynamics. Defining

$$g := \inf_i \inf_\varphi |B_i| \mathrm{gap}_{B_i}^\varphi \, ,$$

the Dirichlet form (3.2.1) gives that, for any function f,

$$\mathrm{Var}_{B_i}^\varphi(f) \leq \frac{\mathcal{E}_{B_i}^\varphi(f)}{\mathrm{gap}_{B_i}^\varphi} \leq \frac{|B_i|}{g} \mathcal{E}_{B_i}^\varphi(f) \, .$$

Combining this with definition (3.2.2) of $\mathcal{E}(\cdot)$,

$$\mathcal{E}_{\mathcal{B}}^\eta(f) = \frac{1}{k} \sum_{\varphi \in \Omega} \mu_W^\eta(\varphi) \sum_i \mathrm{Var}_{B_i}^\varphi(f) \leq \frac{1}{kg} \sum_{\varphi \in \Omega} \mu_W^\eta(\varphi) \sum_i |B_i| \mathcal{E}_{B_i}^\varphi(f) \, .$$

On the other hand, definition (3.2.2) again implies that

$$\sum_{\varphi \in \Omega} \mu_W^\eta(\varphi) \sum_i |B_i| \mathcal{E}_{B_i}^\varphi(f)$$

$$= \sum_{\varphi \in \Omega} \mu_W^\eta(\varphi) \frac{1}{2} \sum_{\sigma \in \Omega} \sum_i \sum_{x \in B_i} \mu_{B_i}^\varphi(\sigma) |B_i| P_{B_i}^\varphi(\sigma, \sigma^x) [f(\sigma) - f(\sigma^x)]^2$$

$$\leq \frac{1}{2} \sup_{x \in W} \#\{i : B_i \ni x\} \sum_{\sigma \in \Omega} \mu_W^\eta(\sigma) \sum_{x \in W} |W| P_W^\sigma(\sigma, \sigma^x) [f(\sigma) - f(\sigma^x)]^2$$

$$= |W| \sup_{x \in W} \#\{i : B_i \ni x\} \mathcal{E}_W^\eta(f) \, ,$$

where σ^x is the configuration obtained from σ by flipping the spin at x, and we used the fact that

$$|B_i| P_{B_i}^\sigma(\sigma, \sigma^x) = |W| P_W^\sigma(\sigma, \sigma^x) \quad \text{for any } i \in [k] \text{ and } x \in B_i \, .$$

Altogether, we obtain that

$$\mathcal{E}_{\mathcal{B}}^\eta(f) \leq \frac{|W|}{kg} \sup_{x \in W} \#\{i : B_i \ni x\} \mathcal{E}_W^\eta(f) \, .$$

Recalling that the single-site dynamics and the block-dynamics have the same stationary measure,

$$\frac{\mathcal{E}_{\mathcal{B}}^{\eta}(f)}{\mathrm{Var}_{W}^{\eta}(f)} = \frac{\mathcal{E}_{\mathcal{B}}^{\eta}(f)}{\mathrm{Var}_{\mathcal{B}}^{\eta}(f)} \geq \mathtt{gap}_{\mathcal{B}}^{\eta}$$

(where we again applied inequality (3.2.1)), thus

$$\frac{\mathcal{E}_{W}^{\eta}(f)}{\mathrm{Var}_{W}^{\eta}(f)} \geq \frac{k}{|W|} g\Big(\sup_{x \in W} \#\{i : B_i \ni x\} \Big)^{-1} \mathtt{gap}_{\mathcal{B}}^{\eta} \;.$$

The proof is now completed by choosing f to be the eigenfunction that corresponds to the second eigenvalue of P_{W}^{η} (achieving $\mathtt{gap}_{W}^{\eta}$), with a final application of (3.2.1). ■

The above proposition can be applied, as part of the spectral gap analysis, to reduce the size of the base graph (though with an arbitrary boundary condition), provided that one can estimate the gap of the corresponding block dynamics chain.

3.2.6 Decomposition of Markov chains

In order to bound the spectral gap of the block dynamics, we require a result of [48], which analyzes the spectral gap of a Markov chain via its decomposition into a *projection* chain and a *restriction* chain.

Consider an ergodic Markov chain on a finite state space Ω with transition kernel $P : \Omega \times \Omega \to [0,1]$ and stationary distribution $\pi : \Omega \to [0,1]$. We assume that the Markov chain is *time-reversible*, that is to say, it satisfies the detailed balance condition

$$\pi(x)P(x,y) = \pi(y)P(y,x) \text{ for all } x,y \in \Omega \;.$$

Let $\Omega = \Omega_0 \cup \ldots \cup \Omega_{m-1}$ be a decomposition of the state space into m disjoint sets. Writing $[m] := \{0, \ldots, m-1\}$, we define $\bar{\pi} : [m] \to [0,1]$ as

$$\bar{\pi}(i) := \pi(\Omega_i) = \sum_{x \in \Omega_i} \pi(x)$$

and define $\bar{P} : [m] \times [m] \to [0,1]$ to be

$$\bar{P}(i,j) := \frac{1}{\bar{\pi}(i)} \sum_{x \in \Omega_i, y \in \Omega_j} \pi(x)P(x,y) \;.$$

The Markov chain on the state space $[m]$ whose transition kernel is $\bar{P}$ is called the *projection* chain, induced by the partition $\Omega_0, \ldots, \Omega_{m-1}$. It is easy to verify that, as the original Markov

chain is reversible with respect to π, the projection chain is reversible with respect to the stationary distribution $\bar{\pi}$.

In addition, each Ω_i induces a *restriction* chain, whose transition kernel $P_i : \Omega_i \times \Omega_i \rightarrow [0, 1]$ is given by

$$P_i(x, y) = \begin{cases} P(x, y), & \text{if } x \neq y, \\ 1 - \sum_{z \in \Omega_i \setminus \{x\}} P(x, z), & \text{if } x = y \ . \end{cases}$$

Again, the restriction chain inherits its reversibility from the original chain, and has a stationary measure π_i, which is simply π restricted to Ω_i:

$$\pi_i(x) := \pi(x)/\bar{\pi}(i) \ \text{ for all } x \in \Omega_i \ .$$

In most applications, the projection chain and the different restriction chains are all irreducible, and thus the various stationary distributions $\bar{\pi}$ and $\pi_0, \ldots, \pi_{m-1}$ are all unique.

The following result provides a lower bound on the spectral gap of the original Markov chain given its above described decomposition:

Theorem 3.2.6 ([48, Theorem 1]). *Let P be the transition kernel of a finite reversible Markov chain, and let* gap *denote its spectral gap. Consider the decomposition of the chain into a projection chain and m restriction chains, and denote their corresponding spectral gaps by* $\bar{\text{gap}}$ *and* $\text{gap}_0, \ldots, \text{gap}_{m-1}$. *Define*

$$\text{gap}_{\min} := \min_{i \in [m]} \text{gap}_i \ , \qquad \gamma := \max_{i \in [m]} \max_{x \in \Omega_i} \sum_{y \in \Omega \setminus \Omega_i} P(x, y) \ .$$

Then gap, *the spectral gap of the original Markov chain, satisfies:*

$$\text{gap} \geq \frac{\bar{\text{gap}}}{3} \ \wedge \ \frac{\bar{\text{gap}} \cdot \text{gap}_{\min}}{3\gamma + \bar{\text{gap}}} \ .$$

The main part of Section 3.4 will be devoted to the analysis of the projection chain, in an effort to bound the spectral gap of our block dynamics via the above theorem.

3.3 Spatial mixing of Ising model on trees

In this section, we will establish a spatial-mixing type result for the Ising model on a general (not necessarily regular) finite tree under an arbitrary boundary condition. This result (namely, Proposition 3.3.1) will later serve as the main ingredient in the proof of Theorem 6 (see Section 3.4). Throughout this section, let $\beta > 0$ be an arbitrary inverse-temperature

and $\theta = \tanh\beta$.

We begin with a few notations. Let T be a tree rooted at ρ with a boundary condition $\tau \in \{\pm 1\}^{\partial T}$ on its leaves, and μ^τ be the corresponding Gibbs measure.

For any $v \in T$, denote by T_v the subtree of T containing v and its all descendants. In addition, for any $B \subset A \subset T$ and $\sigma \in \{\pm 1\}^A$, denote by σ_B the restriction of σ to the sites of B. We then write μ_v^τ for the Gibbs measure on the subtree T_v given the boundary $\tau_{\partial T_v}$.

Consider $\hat{T} \subset T \setminus \partial T$, a subtree of T that contains the root ρ, and write $\hat{T}_v = T_v \cap \hat{T}$. Similar to the above definitions for T, we denote by $\hat{\mu}^\xi$ the Gibbs measure on $\hat{T}$ given the boundary condition $\xi \in \{\pm 1\}^{\partial\hat{T}}$, and let $\hat{\mu}_v^\xi$ be the Gibbs measure on $\hat{T}_v$ given the boundary $\xi_{\partial\hat{T}_v}$.

The following two measures are the conditional distributions of μ_v^τ on the boundary of the subtree $\hat{T}_v$ given the spin at its root v:

$$Q_v^+(\xi) := \mu_v^\tau\big(\sigma_{\partial\hat{T}_v} = \xi_{\partial\hat{T}_v} \mid \sigma(v) = 1\big) \qquad \text{for } \xi \in \{\pm 1\}^{\partial\hat{T}} ,$$
$$Q_v^-(\xi) := \mu_v^\tau\big(\sigma_{\partial\hat{T}_v} = \xi_{\partial\hat{T}_v} \mid \sigma(v) = -1\big) \qquad \text{for } \xi \in \{\pm 1\}^{\partial\hat{T}} .$$

We can now state the main result of this section, which addresses the problem of reconstructing the spin at the root of the tree from its boundary.

Proposition 3.3.1. *Let $\hat{T}$ be as above, let $0 < \theta \leq \frac{3}{4}$ and define*

$$\Delta := \int \hat{\mu}^\xi(\sigma(\rho) = 1)dQ_\rho^+(\xi) - \int \hat{\mu}^\xi(\sigma(\rho) = 1)dQ_\rho^-(\xi) .$$

Then there exists an absolute constant $\kappa > \frac{1}{100}$ such that

$$\Delta \leq \frac{\mathrm{cap}_2(\hat{T})}{\kappa(1 - \theta)} ,$$

where the resistances are assigned to be $R_{(u,v)} = \theta^{-2\mathrm{dist}(\rho,v)}$. Furthermore, this also holds for any external field $h \in \mathbb{R}$ on the root ρ.

To prove the above theorem, we consider the notion of the log-likelihood ratio at a vertex v with respect to $\hat{T}_v$ given the boundary $\xi_{\partial\hat{T}_v}$:

$$x_v^\xi = \log\left(\frac{\hat{\mu}_v^\xi(\sigma(v) = +1)}{\hat{\mu}_v^\xi(\sigma(v) = -1)}\right) , \tag{3.3.1}$$

as well as the following quantity, analogous to Δ (defined in Proposition 3.3.1):

$$m_v := \int x_v^\xi dQ_v^+ - \int x_v^\xi dQ_v^- \ . \tag{3.3.2}$$

As we will later explain, $m_v \geq 0$ for any $v \in T$, and we seek an upper bound on this quantity. One of the main results of [73] was such an estimate for the case of *free* boundary condition, yet in our setting we have an *arbitrary* boundary condition (adding a considerable amount of difficulties to the analysis). The following theorem extends the upper bound on m_ρ to any boundary; to avoid confusion, we formulate this bound in terms of the same absolute constant κ given in Proposition 3.3.1.

Theorem 3.3.2. *Let $\hat{T}$ and m_ρ be as above, and let $0 < \theta \leq \frac{3}{4}$. Then there exists an absolute constant $\kappa > \frac{1}{100}$ such that*

$$m_\rho \leq \frac{\mathrm{cap}_2(\hat{T})}{\kappa(1-\theta)/4} \ ,$$

where the resistances are assigned to be $R_{(u,v)} = \theta^{-2\mathrm{dist}(\rho,v)}$.

Proof of Theorem 3.3.2

As mentioned above, the novelty (and also the main challenge) in the result stated in Theorem 3.3.2 is the presence of the arbitrary boundary condition τ, which eliminates most of the symmetry that one has in the free boundary case. Note that this symmetry was a crucial ingredient in the proof of [73] for the free boundary case (namely, in that case Q_v^+ and Q_v^- are naturally symmetric).

In order to tackle this obstacle, we need to track down the precise influence of the boundary condition τ on each vertex $v \in T$. We then incorporate this information in the recursive analysis that appeared (in a slightly different form) in [62]. This enables us to relate the recursion relation of the m_v-s to that of the L^2-capacity.

The following quantity captures the above mentioned influence of τ on a given vertex $v \in T$:

$$x_v^* = \log\left(\frac{\mu_v^\tau(\sigma(v) = 1)}{\mu_v^\tau(\sigma(v) = -1)}\right) \ . \tag{3.3.3}$$

Notice that x_v^* has a similar form to x_v^ξ (defined in (3.3.1)), and is essentially the log-likelihood ratio at v induced by the boundary condition τ. The quantity x_v^ξ is then the log-likelihood ratio that in addition considers the extra constraints imposed by ξ. Also note that a free boundary condition corresponds to the case where $x_v^* = 0$ for all $v \in T$.

To witness the effect of x_v^*, consider the probabilities of propagating a spin from a parent

v to its child w, formally defined by

$$p_{v,w}^\tau(1,1) := \mu_v^\tau(\sigma(w) = 1 \mid \sigma(v) = 1) \ ,$$
$$p_{v,w}^\tau(1,-1) := \mu_v^\tau(\sigma(w) = -1 \mid \sigma(v) = 1) \ ;$$

we define $p_{v,w}^\tau(-1,1)$ and $p_{v,w}^\tau(-1,-1)$ analogously. The next simple lemma shows the relation between x_v^* and these probabilities.

Lemma 3.3.3. *The following holds for any $(v,w) \in T$:*

$$p_{v,w}^\tau(1,1) - p_{v,w}^\tau(-1,1) = D_w^*\theta \ ,$$

where $D_w^* := (\cosh\beta)^2 / \left(\cosh^2\beta + \cosh^2(x_w^*/2) - 1\right)$.

Proof. Recalling definition (3.3.3) of x_v^*, we can translate the boundary condition τ into an external field $x_w^*/2$ on the vertex w when studying the distribution of its spin. Hence,

$$
\begin{aligned}
p_{v,w}^\tau(1,1) - p_{v,w}^\tau(-1,1) &= \frac{e^{\beta+x_w^*/2}}{e^{\beta+x_w^*/2} + e^{-\beta-x_w^*/2}} - \frac{e^{-\beta+x_w^*/2}}{e^{-\beta+x_w^*/2} + e^{\beta-x_w^*/2}} \\
&= \frac{e^{2\beta} - e^{-2\beta}}{e^{x_w^*} + e^{-x_w^*} + e^{2\beta} + e^{-2\beta}} \\
&= \frac{\cosh^2\beta}{\cosh^2\beta + \cosh^2(x_w^*/2) - 1} \tanh\beta \ ,
\end{aligned}
$$

as required. ∎

Remark 1. In the free boundary case, we have $p_{v,w}(1,1) - p_{v,w}(-1,1) = \theta$. For a boundary condition τ, the coefficient $0 < D_w^* \le 1$ represents the contribution of this boundary to the propagation probability.

We now turn our attention to m_v. As mentioned before, the fact that $m_v \ge 0$ follows from its definition (3.3.2). Indeed, the monotonicity of the Ising model implies that the measure Q_v^+ stochastically dominates the measure Q_v^- (with respect to the natural partial order on the configurations of $\partial\hat{T}_v$). For instance, it is easy to see this by propagating 1 and -1 spins from the root to the bottom, and applying a monotone coupling on these two processes. Finally, x_v^ξ is monotone increasing in ξ (again by the monotonicity of the Ising model), thus $m_v \ge 0$.

The first step in establishing the recursion relation of m_v (that would lead to the desired upper bound) would be to relate m_v to some quantities associated with its children, as stated next.

Lemma 3.3.4. *For any $v \in \hat{T} \setminus \partial\hat{T}$, we have that*

$$m_v = \theta \sum_{w:(v,w)\in\hat{T}} D_w^* \left(\int f(x_w^\xi) dQ_w^+(\xi) - \int f(X_w^\xi) dQ_w^-(\xi) \right) ,$$

where

$$f(x) = \log\left(\frac{\cosh(x/2) + \theta \sinh(x/2)}{\cosh(x/2) - \theta \sinh(x/2)} \right) . \tag{3.3.4}$$

Proof. We need the following well-known lemma, that appeared in [62] in a slightly different form; see also [5] and [73, Lemma 4.1].

Lemma 3.3.5 ([62],[73] (reformulated)). *Let f be as in (3.3.4). For all $v \in \hat{T} \setminus \partial\hat{T}$ and $\xi \in \{\pm1\}^{\partial\hat{T}}$, the following holds: $x_v^\xi = \sum_{w:(v,w)\in\hat{T}} f(x_w^\xi)$.*

According to this lemma, we obtain that

$$m_v = \sum_{w:(v,w)\in\hat{T}} \left(\int f(x_w^\xi) dQ_v^+(\xi) - \int f(x_w^\xi) dQ_v^-(\xi) \right) . \tag{3.3.5}$$

Noting that x_w^ξ is actually a function of $\xi_{\partial\hat{T}_w}$, we get that

$$\int f(x_w^\xi) dQ_v^+(\xi) = \int f(x_w^\xi) d(p_{v,w}^\tau(1,1)Q_w^+(\xi) + p_{v,w}^\tau(1,-1)Q_w^-(\xi)) ,$$

and similarly, we have

$$\int f(x_w^\xi) dQ_v^-(\xi) = \int f(x_w^\xi) d(p_{v,w}^\tau(-1,1)Q_w^+(\xi) + p_{v,w}^\tau(-1,-1)Q_w^-(\xi)) .$$

Combining these two equalities, we deduce that

$$\int f(x_w^\xi) dQ_v^+(\xi) - \int f(x_w^\xi) dQ_v^-(\xi)$$
$$= (p_{v,w}^\tau(1,1) - p_{v,w}^\tau(-1,1)) \left(\int f(x_w^\xi) dQ_w^+(\xi) - \int f(x_w^\xi) dQ_w^-(\xi) \right)$$
$$= \theta D_w^* \left(\int f(x_w^\xi) dQ_w^+(\xi) - \int f(x_w^\xi) dQ_w^-(\xi) \right) , \tag{3.3.6}$$

where in the last inequality we applied Lemma 3.3.3. Plugging (3.3.6) into (3.3.5) now completes the proof of the lemma. ∎

Observe that in the free boundary case, $Q_v^+(\xi) = Q_v^-(-\xi)$ for any ξ. Unfortunately, the

presence of the boundary τ breaks this symmetry, causing the distributions Q_v^+ and Q_v^- to become skewed. Nevertheless, we can still relate these two distributions through the help of x_v^*, as formulated by the following lemma.

Lemma 3.3.6. *For $v \in T$, let Q_v be the following distribution over $\{\pm 1\}^{\partial \hat{T}}$:*

$$Q_v(\xi) = Q_v^\tau(\xi) := \mu_v^\tau \left(\sigma_{\partial \hat{T}_v} = \xi_{\partial \hat{T}_v} \right) \ .$$

Then for all $\xi \in \{\pm 1\}^{\partial \hat{T}}$, we have

$$Q_v^+(\xi) - Q_v^-(\xi) = C_v^* \left(\tanh \frac{x_v^\xi}{2} - \tanh \frac{x_v^*}{2} \right) Q_v(\xi) \ ,$$

where $C_v^* = 2 \cosh^2(x_v^*/2)$.

Proof. It is clear from the definitions of Q_v^+, Q_v^- and Q_v that

$$Q_v^+(\xi) = \frac{Q_v(\xi) \mu_v^\tau(\sigma(v) = 1 \,|\, \xi)}{\mu_v^\tau(\sigma(v) = 1)} = \frac{1 + \tanh(x_v^\xi/2)}{1 + \tanh(x_v^*/2)} Q_v(\xi) \ ,$$

$$Q_v^-(\xi) = \frac{Q_v(\xi) \mu_v^\tau(\sigma(v) = -1 \,|\, \xi)}{\mu_v^\tau(\sigma(v) = -1)} = \frac{1 - \tanh(x_v^\xi/2)}{1 - \tanh(x_v^*/2)} Q_v(\xi) \ .$$

Hence, a straightforward calculation gives that

$$\begin{aligned}
Q_v^+(\xi) - Q_v^-(\xi) &= \frac{2 \left(\tanh(x_v^\xi/2) - \tanh(x_v^*/2) \right)}{(1 + \tanh(x_v^*/2))(1 - \tanh(x_v^*/2))} Q_v(\xi) \\
&= 2 \cosh^2 \left(\frac{x_v^*}{2} \right) \left(\tanh \frac{x_v^\xi}{2} - \tanh \frac{x_v^*}{2} \right) Q_v(\xi) \ ,
\end{aligned}$$

as required. ∎

The following technical lemma will allow us to combine Lemmas 3.3.4, 3.3.6 and obtain an upper bound on m_v in terms of $\{m_w : (v, w) \in \hat{T}\}$. Note that the constant κ mentioned next is in fact the absolute constant κ in the statement of Theorem 3.3.2.

Lemma 3.3.7. *Let f be defined as in (3.3.4) for some $0 < \theta \leq \frac{3}{4}$. Then*

$$|f(x) - f(y)| \leq 2f(|x - y|/2) \text{ for any } x, y \in \mathbb{R} \ , \tag{3.3.7}$$

and there exists a universal constant $\kappa > \frac{1}{100}$ such that for any two constants $C_1, C_2 \geq 1$ with

$C_2 \geq 1 + \left(\frac{1}{2}C_1 - 1\right)\left(1 - \theta^2\right)$ *and any $\delta > 0$ we have*

$$f(\delta)\left(1 + 4\kappa(1 - \theta)C_1\delta \tanh(\delta/2)\right) \leq C_2\theta\delta . \tag{3.3.8}$$

Proof. We first show (3.3.7). Put $\delta = |x - y|$, and define

$$h(\delta) = \sup_t |f(t + \delta) - f(t)| .$$

We claim that

$$h(\delta) = f(\delta/2) - f(-\delta/2) = 2f(\delta/2) . \tag{3.3.9}$$

The second equality follows from the fact that $f(x)$ is an odd function. To establish the first equality, a straightforward calculation gives that

$$f'(x) = \frac{\theta}{1 + (1 - \theta^2)\sinh^2(x/2)} ,$$

and it follows that $f'(x)$ is an even non-negative function which is decreasing in $x \geq 0$. The following simple claim therefore immediately implies (3.3.9):

Claim 3.3.8. *Let $g(t) \geq 0$ be an even function that is decreasing in $t \geq 0$. Then $G(t) = \int_0^t g(x)dx$ has $G(t + \delta) - G(t) \leq 2G(\delta/2)$ for any t and $\delta > 0$.*

Proof. Fix $\delta > 0$ and define $F(t)$ as follows:

$$F(t) = G(t + \delta) - G(t) .$$

We therefore have $F'(t) = g(t + \delta) - g(t)$. Noticing that

$$\begin{cases} |t + \delta| \geq |t| & \text{if } t \geq -\frac{\delta}{2} \\ |t + \delta| \leq |t| & \text{otherwise} \end{cases} ,$$

the assumption on g now gives that $F'(t) \leq 0$ while $t \geq -\frac{\delta}{2}$ and otherwise $F'(t) \geq 0$. Altogether, we deduce that

$$F(t) \leq F\left(-\frac{\delta}{2}\right) = G\left(\frac{\delta}{2}\right) - G\left(-\frac{\delta}{2}\right) = 2G\left(\frac{\delta}{2}\right) ,$$

as required. ∎

It remains to show that (3.3.8) holds for some $\kappa = \kappa(\theta_0) > 0$. Clearly, it suffices to establish this statement for $C_2 = \left[1 + \left(\frac{1}{2}C_1 - 1\right)\left(1 - \theta^2\right)\right] \vee 1$ and any $C_1 \geq 1$. Rearranging

(3.3.8), we are interested in a lower bound on

$$\inf_{\theta \le \theta_0 \,,\, t > 0 \,,\, C_1 \ge 1} \frac{\left[\left(1 + \left(\frac{C_1}{2} - 1\right)\left(1 - \theta^2\right)\right) \vee 1\right]\theta t - f(t)}{4C_1(1 - \theta)tf(t)\tanh(\frac{t}{2})} \,. \tag{3.3.10}$$

First, consider the case $1 \le C_1 < 2$. We then have $C_2 = 1$, and the expression being minimized in (3.3.10) takes the form:

$$\frac{\theta t - f(t)}{4C_1(1 - \theta)tf(t)\tanh(\frac{t}{2})} > \frac{\theta t - f(t)}{8(1 - \theta)tf(t)\tanh(\frac{t}{2})} := \Gamma(t, \theta) \,,$$

where the inequality is by our assumption that $C_1 < 2$. We therefore have that $\inf_{\theta \le \theta_0 \,,\, t > 0} \Gamma(t, \theta)$ minimizes (3.3.10) for $C_1 < 2$, and will next show that this is also the case for $C_1 \ge 2$ under a certain condition. Indeed, letting

$$g(t, \theta, C_1) := \frac{\left[1 + \left(\frac{C_1}{2} - 1\right)\left(1 - \theta^2\right)\right]\theta t - f(t)}{4C_1(1 - \theta)tf(t)\tanh(t/2)} \,,$$

it is easy to verify that the following holds:

$$\frac{\partial g}{\partial C_1} = \frac{f(t) - \theta^3 t}{4C_1^2(1 - \theta)tf(t)\tanh(\frac{t}{2})} \,,$$

hence g is increasing in C_1 for every θ, t such that $f(t) > \theta^3 t$. Therefore,

$$g(t, \theta, C_1) \ge g(t, \theta, 2) = \Gamma(t, \theta) \text{ for all } t, \theta \text{ such that } f(t) > \theta^3 t \,.$$

Before analyzing $\Gamma(t, \theta)$, we will treat the values of θ, t such that $f(t) \le \theta^3 t$. Assume that the case, and notice that the numerator of g then satisfies

$$\left[1 + \left(\frac{C_1}{2} - 1\right)(1 - \theta^2)\right]\theta t - f(t)$$
$$\ge \left[1 + \left(\frac{C_1}{2} - 1\right)(1 - \theta^2) - \theta^2\right]\theta t = \theta(1 - \theta^2)t\frac{C_1}{2} \,,$$

and thereby the dependency on C_1 vanishes:

$$g(t, \theta, C_1) \ge \frac{\theta(1 - \theta^2)t/2}{4(1 - \theta)tf(t)\tanh(t/2)} = \frac{\theta(1 + \theta)}{8f(t)\tanh(t/2)} \,.$$

Since both $\tanh(t/2)$ and $f(t)$ are monotone increasing in t and are bounded from above by

1 and $\log\left(\frac{1+\theta}{1-\theta}\right)$ respectively, we get

$$g(t,\theta,C_1) \geq \frac{\theta(1+\theta)}{8\log\left(\frac{1+\theta}{1-\theta}\right)} \geq \frac{\theta(1+\theta)}{8\frac{2\theta}{1-\theta}} = \frac{1-\theta^2}{16} > \frac{1}{40} \ , \tag{3.3.11}$$

where the second inequality is by the fact that $\log(1+x) \leq x$ for any $x > 0$, and the last inequality follows by the assumption $\theta \leq \frac{3}{4}$.

It thus remains to establish a uniform lower bound on $\Gamma(t,\theta)$. In what follows, our choice of constants was a compromise between simplicity and the quality of the lower bound, and we note that one can easily choose constants that are slightly more optimal.

Assume first that $\theta \geq \theta_0 \geq 0$ for some θ_0 to be defined later. Notice that

$$\tilde{f}(t,\theta) := \frac{1}{\theta}f(t,\theta) = 2\sum_{i=0}^{\infty} \frac{\tanh^{2i+1}(t/2)}{2i+1}\theta^{2i} \ ,$$

and so $\tilde{f}(t,\theta)$ is strictly increasing in θ for any $t > 0$. Since

$$\begin{aligned}
\Gamma(t,\theta) &= \frac{\theta t - f(t)}{8(1-\theta)tf(t)\tanh(\frac{t}{2})} \\
&\geq \frac{\theta t - f(t)}{8(1-\theta_0)tf(t)\tanh(\frac{t}{2})} = \frac{t - \tilde{f}(t)}{8(1-\theta_0)t\tilde{f}(t)\tanh(\frac{t}{2})} \ ,
\end{aligned}$$

we have that Γ is monotone decreasing in θ for any such t, and therefore $\Gamma(t,\theta) \geq \frac{1}{8(1-\theta_0)}\tilde{\Gamma}(t)$, where $\tilde{\Gamma}$ is defined as follows:

$$\tilde{\Gamma}(t) := \frac{\theta t - f(t,\theta)}{tf(t,\theta)\tanh(\frac{t}{2})} \quad \text{with respect to } \theta = \frac{3}{4} \ . \tag{3.3.12}$$

Recall that the Taylor expansion of $f(t,\theta)$ around 0 is $\theta t - \frac{\theta(1-\theta^2)}{12}t^3 + O(t^5)$. It is easy to verify that for $\theta = \frac{3}{4}$ this function satisfies

$$f(t,\theta) \leq \theta t - \frac{(\theta t)^3}{20} \quad \text{for } \theta = \frac{3}{4} \text{ and any } 0 < t \leq 3 \ .$$

Adding the fact that $\tanh(x) \leq x$ for all $x \geq 0$, we immediately obtain that

$$\tilde{\Gamma} \geq \frac{\theta^3 t^3}{20t(\theta t)(t/2)} = \frac{\theta^2}{10} > \frac{1}{20} \quad \text{for all } 0 < t \leq 3 \ .$$

On the other hand, for $t \geq 3$ we can use the uniform upper bounds of 1 and $\log(\frac{1+\theta}{1-\theta})$ for $\tanh(t/2)$ and $f(t)$ respectively, and gain that

$$\tilde{\Gamma} \geq \frac{\theta t - \log(\frac{1+\theta}{1-\theta})}{t \log(\frac{1+\theta}{1-\theta})} = \frac{\theta}{\log(\frac{1+\theta}{1-\theta})} - \frac{1}{t} \geq \frac{1}{20} \quad \text{for all } t \geq 3 .$$

Altogether, as $\Gamma \geq \frac{1}{8(1-\theta_0)}\tilde{\Gamma}$, we can conclude that $\Gamma \geq [160(1-\theta_0)]^{-1}$.

Note that the trivial choice of $\theta_0 = 0$ already provides a uniform lower bound of $\frac{1}{160}$ for Γ (and hence also for κ, as the lower bound in (3.3.11) is only larger). However, this bound can be improved by choosing another θ_0 and treating the case $0 < \theta \leq \theta_0$ separately. To demonstrate this, take for instance $\theta_0 = \frac{1}{2}$. Since the above analysis gave that $\tilde{\Gamma} \geq \frac{1}{20}$ whenever $\theta \leq \frac{3}{4}$, it follows that

$$\Gamma \geq \frac{1}{160(1-\theta_0)} = \frac{1}{80} \quad \text{for all } \tfrac{1}{2} \leq \theta \leq \tfrac{3}{4} .$$

For $\theta \leq \theta_0$, we essentially repeat this analysis of $\tilde{\Gamma}$, only this time the respective value of θ (that is, the maximum value it can attain) is $\frac{1}{2}$. One can thus verify that in that case,

$$f(t, \theta) \leq \theta t - \frac{(\theta t)^3}{6} \quad \text{for } \theta = \tfrac{1}{2} \text{ and any } 0 < t \leq 2.7 ,$$

and the above argument then shows that

$$\tilde{\Gamma} \geq \frac{\theta^2}{3} = \frac{1}{12} \quad \text{for all } 0 < t \leq 2.7 .$$

On the other hand,

$$\tilde{\Gamma} \geq \frac{\theta}{\log(\frac{1+\theta}{1-\theta})} - \frac{1}{t} \geq \frac{1}{12} \quad \text{for all } t \geq 2.7 ,$$

thus for $\theta = \frac{1}{2}$ we have $\tilde{\Gamma} \geq \frac{1}{12}$ for all $t > 0$. This converts into the lower bound $\Gamma \geq \frac{1}{96}$, thus completing the proof with a final value of $\kappa = \frac{1}{96}$. ∎

Remark 2. Note that the only places where we used the fact that $\theta \leq \frac{3}{4}$ are the lower bound on $g(t, \theta, C_1)$ in (3.3.11) and the analysis of $\tilde{\Gamma}$, as defined in (3.3.12). In both cases, we actually only need to have $\theta \leq \theta_1$ for some constant $\theta_1 < 1$, whose precise value might affect the final value of κ.

Using the above lemma, we are now ready to obtain the final ingredient required for the proof of the recursion relation of m_v, as incorporated in Lemma 3.3.9. This lemma provides

a recursive bound on a quantity that resembles m_v, where instead of integrating over x_v^ξ, we integrate over $f(x_v^\xi)$.

Lemma 3.3.9. *Let f and D_v^* be as in (3.3.4) and Lemma 3.3.3 respectively. There exists a universal constant $\kappa > \frac{1}{100}$ so that for $K = \frac{1}{4}(1-\theta)\kappa$ we have*

$$\int f(x_v^\xi)dQ_v^+(\xi) - \int f(x_v^\xi)dQ_v^-(\xi) \leq \frac{\theta m_v}{D_v^*(1+Km_v)} \ .$$

Proof. Clearly,

$$\int f(x_v^\xi)dQ_v^+(\xi) - \int f(x_v^\xi)dQ_v^-(\xi)$$
$$= \int (f(x_v^\xi) - f(x_v^*))dQ_v^+(\xi) - \int (f(x_v^\xi) - f(x_v^*))dQ_v^-(\xi) \ .$$

Applying Lemma 3.3.6, we then obtain that

$$\int f(x_v^\xi)dQ_v^+(\xi) - \int f(x_v^\xi)dQ_v^-(\xi)$$
$$= C_v^* \int (f(x_v^\xi) - f(x_v^*))\left(\tanh \frac{x_v^\xi}{2} - \tanh \frac{x_v^*}{2} \right)dQ_v(\xi) \ ,$$

and similarly,

$$m_v = C_v^* \int (x_v^\xi - x_v^*)\left(\tanh \frac{x_v^\xi}{2} - \tanh \frac{x_v^*}{2} \right)dQ_v(\xi) \ .$$

Let

$$F(x) = (f(x) - f(x_v^*)) \left(\tanh(x/2) - \tanh(x_v^*/2) \right) \ ,$$
$$G(x) = (x - x_v^*) \left(\tanh(x/2) - \tanh(x_v^*/2) \right) \ ,$$

and define Λ to be the probability measure on $\mathbb{R}$ as:

$$\Lambda(x) := Q_v\left(\{\xi : x_v^\xi = x\} \right) \ .$$

According to this definition, we have

$$\int F(x_v^\xi)dQ_v(\xi) = \int F(x)d\Lambda \ , \quad \text{and} \quad \int G(x_v^\xi)dQ_v(\xi) = \int G(x)d\Lambda \ ,$$

and thus, by the above arguments,

$$\int f(x_v^\xi)dQ_v^+(\xi) - \int f(x_v^\xi)dQ_v^-(\xi) = C_v^* \int F(x)d\Lambda ,$$

$$m_v = C_v^* \int G(x)d\Lambda . \tag{3.3.13}$$

Furthermore, notice that by (3.3.7) and the fact that f is odd and increasing for $x \geq 0$,

$$F(x) \leq 2f\Big(\frac{x - x_v^*}{2}\Big)\Big(\tanh\frac{x}{2} - \tanh\frac{x_v^*}{2}\Big) .$$

and so

$$\int f(x_v^\xi)dQ_v^+(\xi) - \int f(x_v^\xi)dQ_v^-(\xi)$$

$$\leq 2C_v^* \int f\Big(\frac{x - x_v^*}{2}\Big)\Big(\tanh\frac{x}{2} - \tanh\frac{x_v^*}{2}\Big)d\Lambda . \tag{3.3.14}$$

In our next argument, we will estimate $\int G(x)d\Lambda$ and $\int G(x)d\Lambda$ according to the behavior of F and G about x_v^*. Assume that $x_v^* \geq 0$, and note that, although the case of $x_v^* \leq 0$ can be treated similarly, we claim that this assumption does not lose generality. Indeed, if $x_v^* < 0$, one can consider the boundary condition of $\tau' = -\tau$, which would give the following by symmetry:

$$x_v^{*\prime} = -x_v^* , \quad X_v'(-\xi) = -x_v^\xi(\xi) , \quad Q_v'(-\xi) = Q_v(\xi) .$$

Therefore, as $f(\cdot)$ and $\tanh(\cdot)$ are both odd functions, we have that $\int F(x)d\Lambda$ and $\int G(x)d\Lambda$ will not change under the modified boundary condition, and yet $x_v^{*\prime} \geq 0$ as required.

Define

$$I^- := (-\infty, x_v^*] \quad , \quad I^+ := [x_v^*, \infty) .$$

First, consider the case where for either $I = I^+$ or $I = I^-$ we have

$$\begin{cases} \int_I F(x)d\Lambda \geq \frac{1}{2}\int F(x)d\Lambda , \\ \int_I G(x)d\Lambda \geq \frac{1}{2}\int F(x)d\Lambda . \end{cases} \tag{3.3.15}$$

In this case, the following holds

$$\Big(\int F(x)d\Lambda\Big)\Big(\int G(x)d\Lambda\Big) \leq 4\Big(\int_I F(x)d\Lambda\Big)\Big(\int_I G(x)d\Lambda\Big)$$

$$\leq 4\int_I F(x)G(x)d\Lambda \leq 4\int F(x)G(x)d\Lambda ,$$

where in the second line we applied the FKG-inequality, using the fact that both F and G are decreasing in I^- and increasing in I^+. The last inequality followed from the fact that F and G are always non-negative. Note that

$$\int F(x)G(x)d\Lambda = \int (f(x) - f(x_v^*))(x - x_v^*)\left(\tanh\frac{x}{2} - \tanh\frac{x_v^*}{2}\right)^2 d\Lambda \ ,$$

and recall that Claim 3.3.8 applied onto $\tanh(x)$ (which indeed has an even non-negative derivative $\cosh^{-2}(x)$ that is decreasing in $x \geq 0$) gives

$$\tanh\frac{x}{2} - \tanh\frac{y}{2} \leq 2\tanh\left(\frac{x-y}{4}\right) \quad \text{for any } x > y \ .$$

Noticing that each of the factors comprising $F(x)G(x)$ has the same sign as that of $(x - x_v^*)$, and combining this with (3.3.7), it thus follows that

$$\left(\int F(x)d\Lambda\right)\left(\int G(x)d\Lambda\right)$$
$$\leq 16\int f\left(\frac{x-x_v^*}{2}\right)(x - x_v^*)\left(\tanh\frac{x}{2} - \tanh\frac{x_v^*}{2}\right)\tanh\left(\frac{x-x_v^*}{4}\right)d\Lambda \ . \tag{3.3.16}$$

Second, consider the case where for I^+ and I^- as above, we have

$$\begin{cases} \int_{I^+} F(x)d\Lambda \geq \frac{1}{2}\int F(x)d\Lambda \ , \\ \int_{I^-} G(x)d\Lambda \geq \frac{1}{2}\int G(x)d\Lambda \ . \end{cases} \tag{3.3.17}$$

The following definitions of $\tilde{F}$ and $\tilde{G}$ thus capture a significant contribution of F and G to $\int F d\Lambda$ and $\int G d\Lambda$ respectively:

$$\begin{cases} \tilde{F}(s) := F(x_v^* + s) \\ \tilde{G}(s) := G(x_v^* - s) \end{cases} \quad \text{for any } s \geq 0 \ , \tag{3.3.18}$$

By further defining the probability measure $\tilde{\Lambda}$ on $[0, \infty)$ to be

$$\tilde{\Lambda}(s) := \Lambda(x_v^* - s)\mathbf{1}_{\{s \neq 0\}} + \Lambda(x_v^* + s) \quad \text{for any } s \geq 0 \ , \tag{3.3.19}$$

we obtain that

$$\int F(x)d\Lambda \le 2 \int_{I^+} F(x)d\Lambda \le 2 \int_0^\infty \tilde{F}(x)d\tilde{\Lambda} \ ,$$

$$\int G(x)d\Lambda \le 2 \int_{I^-} G(x)d\Lambda \le 2 \int_0^\infty \tilde{G}(x)d\tilde{\Lambda} \ .$$

With both $\tilde{F}$ and $\tilde{G}$ being monotone increasing on $[0, \infty)$, applying the FKG-inequality with respect to $\tilde{\Lambda}$ now gives

$$\left(\int F(x)d\Lambda \right) \left(\int G(x)d\Lambda \right) \le 4 \int_0^\infty \tilde{F}(x)\tilde{G}(x)d\tilde{\Lambda}$$

$$= 4 \int_0^\infty (f(x_v^* + s) - f(x_v^*)) \left(\tanh \frac{x_v^* + s}{2} - \tanh \frac{x_v^*}{2} \right)$$

$$\cdot (-s) \left(\tanh \frac{x_v^* - s}{2} - \tanh \frac{x_v^*}{2} \right) d\tilde{\Lambda} \ .$$

Returning to the measure Λ, the last expression takes the form

$$4 \int_{I^+} (f(x) - f(x_v^*)) \left(\tanh \frac{x}{2} - \tanh \frac{x_v^*}{2} \right)$$

$$\cdot (x - x_v^*) \left(\tanh \frac{x_v^*}{2} - \tanh \frac{2x_v^* - x}{2} \right) d\Lambda$$

$$+ 4 \int_{I^-} (f(2x_v^* - x) - f(x_v^*)) \left(\tanh \frac{2x_v^* - x}{2} - \tanh \frac{x_v^*}{2} \right)$$

$$\cdot (x - x_v^*) \left(\tanh \frac{x}{2} - \tanh \frac{x_v^*}{2} \right) d\Lambda \ .$$

We now apply (3.3.7) and Claim 3.3.8 (while leaving the term $(\tanh \frac{x}{2} - \tanh \frac{x_v^*}{2})$ unchanged in both integrals) to obtain that

$$\left(\int F(x)d\Lambda \right) \left(\int G(x)d\Lambda \right)$$

$$\le 16 \int f\left(\frac{x - x_v^*}{2} \right) \left(\tanh \frac{x}{2} - \tanh \frac{x_v^*}{2} \right) (x - x_v^*) \tanh \left(\frac{x - x_v^*}{4} \right) d\Lambda \ .$$

That is, we have obtained the same bound as in (3.3.16).

It remains to deal with the third case where for I^+ and I^- as above,

$$\begin{cases} \int_{I^-} F(x)d\Lambda \geq \frac{1}{2} \int F(x)d\Lambda \ , \\ \int_{I^+} G(x)d\Lambda \geq \frac{1}{2} \int G(x)d\Lambda \ . \end{cases} \tag{3.3.20}$$

In this case, we modify the definition (3.3.18) of $\tilde{F}$ and $\tilde{G}$ appropriately:

$$\begin{cases} \tilde{F}(s) := F(x_v^* + s) \\ \tilde{G}(s) := G(x_v^* - s) \end{cases} \quad \text{for any } s \geq 0 \ ,$$

and let $\tilde{\Lambda}$ remain the same, as given in (3.3.19). It then follows that

$$\int F(x)d\Lambda \leq 2 \int_{I^-} F(x)d\Lambda \leq 2 \int_0^\infty \tilde{F}(x)d\tilde{\Lambda} \ ,$$

$$\int G(x)d\Lambda \leq 2 \int_{I^+} G(x)d\Lambda \leq 2 \int_0^\infty \tilde{G}(x)d\tilde{\Lambda} \ ,$$

with $\tilde{F}$ and $\tilde{G}$ monotone increasing on $[0, \infty)$. By the FKG-inequality,

$$\left(\int F(x)d\Lambda \right) \left(\int G(x)d\Lambda \right) \leq 4 \int_0^\infty \tilde{F}(x)\tilde{G}(x)d\tilde{\Lambda}$$

$$= 4 \int_0^\infty \left(f(x_v^* - s) - f(x_v^*) \right) \left(\tanh \frac{x_v^* - s}{2} - \tanh \frac{x_v^*}{2} \right)$$

$$\cdot s \left(\tanh \frac{x_v^* + s}{2} - \tanh \frac{x_v^*}{2} \right) d\tilde{\Lambda} \ . \tag{3.3.21}$$

As before, we now switch back to Λ and infer from (3.3.7) and Claim 3.3.8 that

$$\left(\int F(x)d\Lambda \right) \left(\int G(x)d\Lambda \right)$$

$$\leq 16 \int f\left(\frac{x - x_v^*}{2} \right) \left(\tanh \frac{x}{2} - \tanh \frac{x_v^*}{2} \right) (x - x_v^*) \tanh \left(\frac{x - x_v^*}{4} \right) d\Lambda \ ,$$

that is, (3.3.16) holds for each of the 3 possible cases (3.3.15), (3.3.17) and (3.3.20).

Altogether, this implies that

$$
\left(\int f(x_v^\xi) dQ_v^+(\xi) - \int f(x_v^\xi) dQ_v^-(\xi) \right) m_v
$$
$$
= (C_v^*)^2 \left(\int F(x) d\Lambda \right) \left(\int G(x) d\Lambda \right)
$$
$$
\leq 16 C_v^{*2} \int f\left(\frac{x - x_v^*}{2}\right) \left(\tanh \frac{x}{2} - \tanh \frac{x_v^*}{2} \right) (x - x_v^*) \tanh \left(\frac{x - x_v^*}{4} \right) d\Lambda \ .
$$

Therefore, recalling (3.3.14) and choosing $K = \frac{1}{4}(1 - \theta)\kappa$, where κ is as given in Lemma 3.3.7, we have

$$
\left(\int f(x_v^\xi) dQ_v^+(\xi) - \int f(x_v^\xi) dQ_v^-(\xi) \right)(1 + K m_v)
$$
$$
\leq 2 C_v^* \int f\left(\frac{x - x_v^*}{2}\right) \left[1 + 4\kappa(1 - \theta) C_v^* \frac{x - x_v^*}{2} \tanh \frac{x - x_v^*}{4} \right]
$$
$$
\cdot \left(\tanh \frac{x}{2} - \tanh \frac{x_v^*}{2} \right) d\Lambda
$$
$$
\leq 2 C_v^* \int (1/D_v^*) \theta \frac{x - x_v^*}{2} \left(\tanh \frac{x}{2} - \tanh \frac{x_v^*}{2} \right) d\Lambda = \theta \frac{C_v^*}{D_v^*} \int G(x) d\Lambda \ .
$$

where the inequality in the last line is by Lemma 3.3.7 for $\delta = |x - x_v^*|/2$ (the case $x < x_v^*$ follows once again from the fact that f is odd) and a choice of $C_1 = C_v^* = 2\cosh^2(x_v^*/2) \geq 2$ and $C_2 = (1/D_v^*)$ (recall that, by definition, $1/D_v^* = 1 + (\frac{1}{2}C_v^* - 1)(1 - \theta^2) \geq 1$, satisfying the requirements of the lemma). Therefore, (3.3.13) now implies that

$$
\int f(x_v^\xi) dQ_v^+(\xi) - \int f(x_v^\xi) dQ_v^-(\xi) \leq \frac{\theta m_v}{D_v^*(1 + K m_v)} \ ,
$$

as required. ∎

Combining Lemmas 3.3.4 and 3.3.9, we deduce that there exists a universal constant $\kappa > 0$ such that

$$
m_v \leq \sum_{w:(v,w)\in\hat{T}} \frac{\theta^2 m_w}{1 + \frac{1}{4}\kappa(1 - \theta) m_w} \ . \tag{3.3.22}
$$

The proof will now follow from a theorem of [73], that links a function on the vertices of a tree T with its L^2-capacity according to certain resistances.

Theorem 3.3.10 ([73, Theorem 3.2] (reformulated))**.** *Let T be a finite tree, and suppose that there exists some $K > 0$ and positive constants $\{a_v : v \in T\}$ such that for every $v \in T$*

and $x \geq 0$,

$$g_v(x) \leq a_v x / (1 + Kx) \ .$$

Then any solution to the system $x_v = \sum_{w:(v,w) \in T} g_w(x_w)$ satisfies

$$x_\rho \leq \mathrm{cap}_2(T) \big/ K \ ,$$

where the resistances are given by $R_{(u,v)} = \prod_{(x,y) \in \mathcal{P}(\rho,v)} a_y^{-1}$, with $\mathcal{P}(\rho, v)$ denoting the simple path between ρ and v.

Together with inequality (3.3.22), the above theorem immediately gives

$$m_\rho \leq \frac{\mathrm{cap}_2(\hat{T})}{\kappa(1-\theta)/4} \ ,$$

completing the proof of Theorem 3.3.2.   ∎

Proof of Proposition 3.3.1

In order to obtain the required result from Theorem 3.3.2, recall the definition of x_v^ξ for $v \in T$, according to which we can write

$$\hat{\mu}^\xi(\sigma(\rho) = 1) = \big(1 + \tanh(x_\rho^\xi/2 + h)\big) /2 \ ,$$

where h is the mentioned external field at the root ρ. By monotone coupling, we can construct a probability measure Q_c on the space $\{(\xi, \xi') : \xi \geq \xi'\}$ such that the two marginal distributions correspond to Q_ρ^+ and Q_ρ^- respectively. It therefore follows that

$$
\begin{aligned}
\Delta &= \int \Big(\hat{\mu}^\xi(\sigma(\rho) = 1) - \hat{\mu}^{\xi'}(\sigma(\rho) = 1) \Big) dQ_c \\
&= \frac{1}{2} \int \Big(\tanh(x_\rho^\xi/2 + h) - \tanh(x_\rho^{\xi'}/2 + h) \Big) dQ_c \\
&\leq \frac{1}{2} \int \frac{x_\rho^\xi - x_\rho^{\xi'}}{2} dQ_c = \frac{1}{4} m_\rho \leq \frac{\mathrm{cap}_2(\hat{T})}{\kappa(1-\theta)} \ ,
\end{aligned}
$$

where the last inequality follows from Theorem 3.3.2 using the same value of $\kappa \geq \frac{1}{100}$. This completes the proof.   ∎

3.4 Upper bound on the inverse-gap and mixing time

This section is devoted to the proof of the main theorem, Theorem 6, from which it follows that the mixing time of the continuous-time Glauber dynamics for the Ising model on a b-ary tree (with any boundary condition) is poly-logarithmic in the tree size.

Recalling the log-Sobolev results described in Section 5.1, it suffices to show an upper bound of $O(n \log^M n)$ on inverse-gap of the discrete-time chain (equivalently, a lower bound on its `gap`), which would then imply an upper bound of $O(n \log^{M+2} n)$ for the L^2 mixing-time (and hence also for the total-variation mixing-time).

The proof comprises several elements, and notably, uses a block dynamics in order to obtain the required upper bound inductively. Namely, we partition a tree on n vertices to blocks of size roughly $n^{1-\alpha}$ each, for some small $\alpha > 0$, and use an induction hypothesis that treats the worst case boundary condition. The main effort is then to establish a lower bound on the spectral-gap of the block dynamics (as opposed to each of its individual blocks). This is achieved by Theorem 3.4.1 (stated later), whose proof hinges on the spatial-mixing result of Section 3.3, combined with the Markov chain decomposition method.

Throughout this section, let $b \geq 2$ be fixed, denote by $\beta_c = \operatorname{arctanh}(1/\sqrt{b})$ the critical inverse-temperature and let $\theta = \tanh \beta_c$.

3.4.1 Block dynamics for the tree

In what follows, we describe our choice of blocks for the above mentioned block dynamics. Let h denote the height of our b-ary tree (that is, there are b^h leaves in the tree), and define

$$\ell := \alpha h \quad , \quad r := h - \ell , \tag{3.4.1}$$

where $0 < \alpha < \frac{1}{2}$ is some (small) constant to be selected later.

For any $v \in T$, let $B(v, k)$ be the subtree of height $k - 1$ rooted at v, that is, $B(v, k)$ consists of k levels (except when v is less than k levels away from the bottom of T). We further let H_k denote the k-th level of the tree T, that according to this notation contains b^k vertices.

Next, define the set of blocks $\mathcal{B}$ as:

$$\mathcal{B} := \{B(v, r) \ : \ v \in H_\ell \cup \{\rho\}\} \quad \text{for } \ell, r \text{ as above.} \tag{3.4.2}$$

That is, each block is a b-ary tree with r levels, where one of these blocks is rooted at ρ, and will be referred to as the *distinguished block*, whereas the others are rooted at the vertices of H_ℓ.
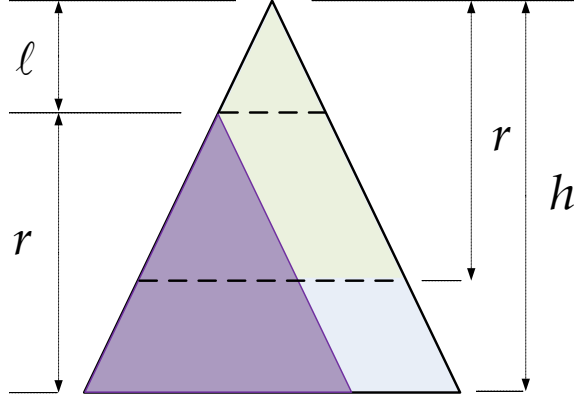
Figure 3.1: Block dynamics for the Ising model on the tree: illustration shows the distinguished block $B(\rho, r)$ as well as a representative block of the form $B(v, r)$ for $v \in H_\ell$.

The following theorem establishes a lower bound on the spectral gap of the above-specified block dynamics (with blocks $\mathcal{B}$).

Theorem 3.4.1. *Consider the Ising model on the b-ary tree at the critical inverse-temperature β_c and with an arbitrary boundary τ. Let $\mathtt{gap}_{\mathcal{B}}^\tau$ be the spectral gap of the corresponding block dynamics with blocks $\mathcal{B}$ as in* (3.4.2)*. The following then holds:*

$$\mathtt{gap}_{\mathcal{B}}^\tau \geq \frac{1}{4(b^\ell + 1)} \left(1 - \frac{\alpha}{\kappa(1-\theta)(1-2\alpha)} \right) ,$$

where $\kappa > 0$ is the absolute constant given in Theorem 3.3.2.

Given the above theorem, we can now derive a proof for the main result.

3.4.2 Proof of Theorem 6

By definition, as $b \geq 2$, we have that

$$\theta = \tanh \beta_c = \frac{1}{\sqrt{b}} \leq \frac{1}{\sqrt{2}} ,$$

hence we can readily choose an absolute constant $0 < \alpha < 1$ such that

$$c(\alpha) := \frac{1}{8} \left(1 - \frac{\alpha}{\kappa(1-\theta)(1-2\alpha)} \right) > 0 .$$

Let $n_h = \sum_{j=0}^{h-1} b^j$ be the number of vertices in a b-ary tree of height h excluding its leaves, and let $\mathtt{gap}_h^\tau$ be the spectral gap of the (single-site) discrete-time Glauber dynamics for the Ising model on a b-ary tree of height h and boundary τ (in the special case of a free boundary condition, n_h should instead include the leaves). Further define

$$g_h = n_h \min_\tau \mathtt{gap}_h^\tau .$$

Recalling the definition of $\mathcal{B}$ according to the above choice of α, we have that each of its blocks is a tree of height $r = (1 - \alpha)h$, and that

$$\sup_{v \in T} \#\{B \in \mathcal{B} \ : \ x \in B\} = 2 ,$$

as each of the vertices in levels $\ell, \ell+1, \ldots, r$ is covered precisely twice in $\mathcal{B}$, while every other vertex is covered precisely once.

Hence, by Proposition 3.2.5 and Theorem 3.4.1, it now follows that for any $h \geq 1/\alpha$ (such that our choices of ℓ, r in (3.4.1) are both non-zero) we have

$$g_h \geq \left(\frac{1}{4(b^\ell + 1)} \left(1 - \frac{\alpha}{\kappa(1-\theta)(1-2\alpha)} \right) \right) g_r \cdot \frac{1}{2} = c(\alpha) g_{(1-\alpha)h} .$$

Having established the induction step, we now observe that, as α is constant, clearly $g_k \geq c'$ holds for any $k \leq 1/\alpha$ and some fixed $c' = c'(\alpha) > 0$. Hence,

$$g_h \geq c' \big(c(\alpha) \big)^{\log_{1-\alpha}(1/h)} = c' h^{-\log\left(\frac{1}{c(\alpha)}\right)/\log\left(\frac{1}{1-\alpha}\right)} ,$$

that is, there exists an absolute constant M (affected by our choice of the absolute constants κ, α) so that the inverse-gap of the continuous-time dynamics with an arbitrary boundary condition τ is at most $g_h^{-1} = O(h^M)$, as required. ∎

3.4.3 Proof of Theorem 3.4.1

In order to obtain the desired lower bound on the spectral gap of the block dynamics, we will apply the method of decomposition of Markov chains, described in Subsection 3.2.6. To this end, we will partition our configuration according to the spins of the subset

$$S := B(\rho, \ell - 1) .$$

Note that S is strictly contained in the distinguished block $B(\rho, r)$, and does not intersect any other $B \in \mathcal{B}$. For $\eta \in \{\pm 1\}^S$, denote the set of configurations which agree with η by

$$\Omega_\eta := \{\sigma \in \Omega : \sigma_S = \eta\} \ .$$

Following the definitions in Subsection 3.2.6, we can now naturally decompose the block dynamics into a projection chain $\bar{P}$ on $\{\pm 1\}^S$ and restriction chains P_η on Ω_η for each $\eta \in \{\pm 1\}^S$. With Theorem 3.2.6 in mind, we now need to provide suitable lower bounds on $\bar{\mathtt{gap}}^\tau$ and $\mathtt{gap}_\eta^\tau$, the respective spectral gaps of $\bar{P}$ and P_η given the boundary condition τ.

We begin with the lower bound on the restriction chain $\mathtt{gap}_\eta^\tau$, formulated in the next lemma.

Lemma 3.4.2. *For any boundary τ and $\eta \in \{\pm 1\}^S$, the spectral gap of the restriction chain P_η on the space Ω_η satisfies $\mathtt{gap}_\eta^\tau \geq 1/(b^\ell + 1)$.*

Proof. Recall that the restriction chain P_η moves from $\sigma \in \Omega_\eta$ to $\sigma' \in \Omega_\eta$ (that is, σ and σ' both agree with η on S) according to the original law of the chain, and remains at σ instead of moving to any $\sigma' \notin \Omega_\eta$. By definition of our block dynamics, this means that with probability $b^\ell/(b^\ell + 1)$ we apply a transition kernel Q_1, that selects one of the blocks rooted at H_ℓ to be updated according to its usual law (since S and all of these blocks are pairwise disjoint). On the other hand, with probability $1/(b^\ell+1)$, we apply a transition kernel Q_2 that updates the distinguished block, yet only allows updates that keep S unchanged (otherwise, the chain remains in place).

We next claim that the update of the distinguished block can only increase the value of $\mathtt{gap}_\eta^\tau$. To see this, consider the chain P_η', in which the distinguished block is never updated; that is, Q_2 described above is replaced by the identity. Clearly, since each of the vertices of $T \setminus S$ appears in (precisely) one of the non-distinguished blocks, the stationary distribution of P_η' is again $\mu_{\tau;\eta}$, the Gibbs distribution with boundary conditions η and τ. Therefore, recalling the Dirichlet form (3.2.2), for any f we clearly have

$$\begin{aligned}
\mathcal{E}_{P_\eta'}(f) &= \frac{1}{2} \sum_{x,y \in \Omega_\eta} [f(x) - f(y)]^2 \, \mu^{\tau;\eta}(x) P_\eta'(x, y) \\
&\leq \frac{1}{2} \sum_{x,y \in \Omega_\eta} [f(x) - f(y)]^2 \, \mu^{\tau;\eta}(x) P_\eta(x, y) = \mathcal{E}_{P_\eta}(f) \ ,
\end{aligned}$$

and thus, by the spectral gap bound in terms of the Dirichlet form (3.2.1),

$$\mathtt{gap}(P_\eta) \geq \mathtt{gap}(P_\eta') \ . \tag{3.4.3}$$

It remains to analyze the chain P'_η, which is in fact a *product chain*, and as such its eigenvalues can be directly expressed in terms of the eigenvalues of its component chains. This well known fact is stated in the following straightforward claim (cf., e.g., [3, Chapter 4] and [54, Lemma 12.11]); we include its proof for completeness.

Claim 3.4.3. *For $j \in [d]$, let P_j be a transition kernel on Ω_j with eigenvalues Λ_j. Let ν be a probability distribution on $[d]$, and define P', the transition matrix of the product chain of the P_j-s on $\Omega' = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_d$, by*

$$P'\big((x_1, \ldots, x_d), (y_1, \ldots, y_d)\big) = \sum_{j=1}^d \nu(j) P_j(x_j, y_j) \prod_{i:i \neq j} \mathbf{1}_{\{x_i = y_i\}} .$$

Then P' has eigenvalues $\left\{ \sum_{j=1}^d \nu(j)\lambda_j : \lambda_j \in \Lambda_j \right\}$ (with multiplicities).

Proof. Clearly, by induction it suffices to prove the lemma for $d = 2$. In this case, it is easy to verify that the transition kernel $\tilde{P}$ can be written as

$$\tilde{P} = \nu(1)(P_1 \otimes I_{\Omega_2}) + \nu(2)(I_{\Omega_1} \otimes P_2) ,$$

where $\otimes$ denotes the matrix tensor-product. Thus, by tensor arithmetic, for any u, v, eigenvectors of P_1, P_2 with corresponding eigenvalues λ_1, λ_2 respectively, $(u \otimes v)$ is an eigenvector of $\tilde{P}$ with a corresponding eigenvalue of $\nu(1)\lambda_1 + \nu(2)\lambda_2$, as required. ∎

In our setting, first notice that Q_1 itself is a product chain, whose components are the b^ℓ chains, uniformly selected, updating each of the non-distinguished blocks. By definition, a single block-update replaces the contents of the block with a sample according to the stationary distribution conditioned on its boundary. Therefore, each of the above mentioned component chains has a single eigenvalue of 1 whereas all its other eigenvalues are 0.

It thus follows that P'_η (a lazy version of Q_1) is another product chain, giving Q_1 probability $b^\ell/(b^\ell + 1)$ and the identity chain probability $1/(b^\ell + 1)$. By Claim 3.4.3, we conclude that the possible eigenvalues of P'_η are precisely

$$\left\{ \frac{1}{b^\ell + 1} + \frac{1}{b^\ell + 1}\sum_{j=1}^{b^\ell} \lambda_j : \lambda_j \in \{0, 1\} \right\} .$$

In particular, $\mathtt{gap}(P'_\eta) = 1/(b^\ell + 1)$, and (3.4.3) now completes the proof. ∎

It remains to provide a bound on $\mathtt{g\bar{a}p}^\tau$, the spectral gap of the projection chain in the decomposition of the block dynamics according to S. This is the main part of our proof of

the lower bound for the spectral gap of the block dynamics, on which the entire proof of Theorem 6 hinges. To obtain this bound, we relate the projection chain to the spatial-mixing properties of the critical Ising model on the tree under various boundary conditions, studied in Section 3.3.

Lemma 3.4.4. *For any boundary τ, the spectral gap of the projection chain $\bar{P}$ on the space $\{\pm 1\}^S$ satisfies*

$$\mathtt{g\bar{a}p}^\tau \geq \frac{1}{b^\ell + 1}\left(1 - \frac{\alpha}{\kappa(1-\theta)(1-2\alpha)}\right) \; ,$$

where $\kappa > 0$ is the absolute constant given in Proposition 3.3.1.

We prove this lemma by establishing a certain contraction property of the projection chain $\bar{P}$. Recall that $\bar{P}(\eta, \eta')$, for $\eta, \eta' \in \{\pm 1\}^S$, is the probability that completing η into a state σ according to the stationary distribution (with boundary η and τ) and then applying the block dynamics transition, gives some σ' that agrees with η' on S.

Let $S^* = H_{\ell-1}$ denote the bottom level of S, and notice that in the above definition of the transition kernel of $\bar{P}$, the value of the spins in $S \setminus S^*$ do not affect the transition probabilities. Therefore, the projection of the chain $\bar{P}$ onto S^* is itself a Markov chain, which we denote by $\bar{P}^*$. In fact, we claim that the eigenvalues of $\bar{P}$ and those of $\bar{P}^*$ are precisely the same (with the exception of additional 0-eigenvalues in $\bar{P}$). To see this, first notice that the eigenfunctions of $\bar{P}^*$ can be naturally extended into eigenfunctions of $\bar{P}$ with the same eigenvalues (as $\bar{P}^*$ is a projection of $\bar{P}$). Furthermore, whenever $\eta_1 \neq \eta_2 \in S$ agree on S^*, they have the same transition probabilities to any $\eta' \in S$, thus contributing a 0-eigenvalue to $\bar{P}$. It is then easy to see that all other eigenvalues of $\bar{P}$ (beyond those that originated from $\bar{P}^*$) must be 0. Altogether,

$$\mathtt{gap}(\bar{P}^*) = \mathtt{gap}(\bar{P}) \quad (= \mathtt{g\bar{a}p}^\tau) \; , \tag{3.4.4}$$

and it remains to give a lower bound for $\mathtt{gap}(\bar{P}^*)$. The next lemma shows that $\bar{P}^*$ is contracting with respect to Hamming distance on $\{\pm 1\}^{S^*}$.

Lemma 3.4.5. *Let $\bar{X}_t^*$ and $\bar{Y}_t^*$ be instances of the chain $\bar{P}^*$, starting from φ and ψ respectively. Then there exists a coupling such that*

$$\mathbf{E}_{\varphi,\psi}\mathrm{dist}(\bar{X}_1^*, \bar{Y}_1^*) \leq \left(\frac{b^\ell}{b^\ell + 1} + \frac{1}{b^\ell + 1} \cdot \frac{1 + (b-1)\ell}{b\kappa(1-\theta)(r-\ell)}\right)\mathrm{dist}(\varphi, \psi) \; .$$

Proof. Clearly, if $\varphi = \psi$ the lemma trivially holds via the identity coupling. In order to understand the setting when $\varphi \neq \psi$, recall the definition of the chain $\bar{P}^*$, which has the following two possible types of moves E_1 and E_2:

1. With probability $1 - \frac{1}{b^\ell+1}$, the block dynamics updates one of the non-distinguished blocks: denote this event by E_1. Since this operation does not affect the value of the spins in the subset S (and in particular, in S^*), the projection chain $\bar{P}$ remains in place in this case (and so does $\bar{P}^*$).

2. With probability $\frac{1}{b^\ell+1}$, the distinguished block is being updated: denote this event by E_2. By the discussion above, this is equivalent to the following. Let η denote the current state of the chain $\bar{P}^*$. First, $T \setminus S$ is assigned values according to the stationary distribution with boundary η and τ. Then, the distinguished block $B(\rho, r)$ is updated given all other spins in the tree, and the resulting value of S (and hence also of S^*) is determined by the new state of the projection chain.

By the triangle inequality, it suffices to consider the case of $\mathrm{dist}(\varphi, \psi) = 1$. Suppose therefore that φ and ψ agree everywhere on S^* except at some vertex ϱ, and that without loss of generality,

$$\varphi(\varrho) = 1 \ , \ \psi(\varrho) = -1 \ .$$

Crucially, the above mentioned procedure for the event E_2 is precisely captured by the spatial-mixing properties that were studied in Section 3.3. Namely, a spin of some site $v \in S^*$ is propagated down the tree T_v (with boundary condition τ), and then the new value of S^* is reconstructed from level $r + 1$, the external boundary of $B(\rho, r)$. We construct a monotone coupling that will accomplish the required contraction property.

First, when propagating the sites $v \in S^*$ with $v \neq \varrho$, we use the identity coupling (recall that $\varphi(v) = \psi(v)$ for all $v \neq \varrho$). Second, consider the process that the spin at ϱ undergoes. For φ, a positive spin is propagated to T_ϱ (with boundary condition τ) and then reconstructed from level $r + 1$ in the tree T (which corresponds to level $r - \ell + 1$ in the subtree T_ϱ), with an additional boundary condition from $T \setminus T_\varrho$ that translates into some external field. For ψ, a negative spin is propagated analogously, and notice that in its reconstruction, the exact same external field applies (as $T \setminus T_\varrho$ was guaranteed to be the same for φ and ψ).

Therefore, applying Proposition 3.3.1 on the tree T_ϱ with respect to the subtree $\hat{T} = B(\varrho, r - \ell + 1)$, we can deduce that

$$\mathbf{E}_{\varphi, \psi}\left(\bar{X}_1^*(\varrho) - \bar{Y}_1^*(\varrho) \,\middle|\, E_2 \right) \leq \frac{\mathrm{cap}_2(B(\varrho, r - \ell + 1))}{\kappa(1 - \theta)} \ , \tag{3.4.5}$$

where $\kappa > \frac{1}{100}$, and the resistances are assigned as

$$R_{(u,v)} = (\tanh \beta_c)^{-2\mathrm{dist}(\varrho, v)} \ .$$

We now turn to estimating the L^2-capacity, which is equivalent to the effective conductance

between ϱ and $\partial B(\varrho, r - \ell + 1)$. This will follow from the well-known Nash-Williams Criterion (cf., e.g., [63]). Here and in what follows, $R_{\text{eff}} := 1/C_{\text{eff}}$ denotes the effective resistance.

Lemma 3.4.6 (Nash-Williams Criterion [72]). *If $\{\Pi_j\}_{j=1}^J$ is a sequence of pairwise disjoint cutsets in a network G that separate a vertex v from some set A, then*

$$R_{\text{eff}}(v \leftrightarrow A) \geq \sum_j \Big(\sum_{e \in \Pi_j} \frac{1}{R_e} \Big)^{-1} .$$

In our case, G is the b-ary tree $B(\varrho, r - \ell + 1)$, and it is natural to select its different levels as the cutsets Π_j. It then follows that

$$R_{\text{eff}}\big(\varrho \leftrightarrow \partial B(\varrho, r - \ell + 1)\big) \geq \sum_{k=1}^{r-\ell+1} (b^k \theta^{2k})^{-1} = r - \ell + 1 , \tag{3.4.6}$$

where we used the fact that $\tanh \beta_c = \theta = 1/\sqrt{b}$. It therefore follows that

$$\text{cap}_2 \left(B(\varrho, r - \ell + 1) \right) \leq \frac{1}{r - \ell} ,$$

which, together with (3.4.5), implies that

$$\mathbf{E}_{\varphi, \psi} \Big(\bar{X}_1^*(\varrho) - \bar{Y}_1^*(\varrho) \,\big|\, E_2 \Big) \leq \frac{1}{\kappa(1 - \theta)(r - \ell)} . \tag{3.4.7}$$

Unfortunately, asides from controlling the probability that the spin at ϱ will coalesce in φ and ψ, we must also consider the probability that ϱ would remain different, and that this difference might be propagated to other vertices in S^* (as part of the update of $B(\rho, r)$). Assume therefore that the we updated the spin at ϱ and indeed $\bar{X}_1^*(\varrho) \neq \bar{Y}_1^*(\varrho)$, and next move on to updating the remaining vertices of S^*. Since our propagation processes corresponding to $\bar{X}^*$ and $\bar{Y}^*$ gave every vertex in $T \setminus T_\varrho$ the same spin, it follows that each vertex $v \in S^*$, $v \neq \varrho$, has the same external field in $\bar{X}^*$ and $\bar{Y}^*$, with the exception of the effect of the spin at ϱ.

We may therefore apply the next lemma of [8], which guarantees that we can ignore this mentioned common external field when bounding the probability of propagating the difference in ϱ.

Lemma 3.4.7 ([8, Lemma 4.1]). *Let T be a finite tree and let $v \neq w$ be vertices in T. Let $\{J_e \geq 0 : e \in E(T)\}$ be the interactions on T, and let $\{H(u) \in \mathbb{R} : u \in V(T)\}$ be an external*

field on the vertices of T. We consider the following conditional Gibbs measures:

$\mu^{+,H}$: *the Gibbs measure with external field H conditioned on $\sigma(v) = 1$.*

$\mu^{-,H}$: *the Gibbs measure with external field H conditioned on $\sigma(v) = -1$.*

Then $\mu^{+,H}(\sigma(w)) - \mu^{-,H}(\sigma(w))$ achieves its maximum at $H \equiv 0$.

In light of the discussion above, Lemma 3.4.7 gives that

$$\mathbf{E}_{\varphi,\psi}\Big(\frac{1}{2}\sum_{v\in S^*}(\bar{X}_1^*(v) - \bar{Y}_1^*(v)) \,\big|\, E_2\Big)$$

$$\leq \mathbf{E}_{\varphi,\psi}\Big(\bar{X}_1^*(\varrho) - \bar{Y}_1^*(\varrho) \,\big|\, E_2\Big)\Big(1 + \sum_{k=1}^{\ell-1}\frac{b-1}{b}b^k\theta^{2k}\Big)$$

$$\leq \frac{1 + (b-1)(\ell-1)/b}{\kappa(1-\theta)(r-\ell)} = \frac{1 + (b-1)\ell}{b\kappa(1-\theta)(r-\ell)} \,,$$

where in the first inequality we used the propagation property of the Ising model on the tree (Claim 3.2.1), and in the second one we used the fact that $\theta = \tanh(\beta_c) = 1/\sqrt{b}$, as well as the estimate in (3.4.7).

We conclude that there exists a monotone coupling of $\bar{X}_t^*$ and $\bar{Y}_t^*$ with

$$\mathbf{E}_{\varphi,\psi}\Big(\mathrm{dist}(\bar{X}_1^*, \bar{Y}_1^*) \,\big|\, E_2\Big) \leq \frac{1 + (b-1)\ell}{b\kappa(1-\theta)(r-\ell)} \,,$$

which then directly gives that

$$\mathbf{E}_{\varphi,\psi}\big(\mathrm{dist}(\bar{X}_1^*, \bar{Y}_1^*)\big) \leq \frac{b^\ell}{b^\ell + 1} + \frac{1}{b^\ell + 1} \cdot \frac{1 + (b-1)\ell}{b\kappa(1-\theta)(r-\ell)} \,,$$

as required. ∎

The above contraction property will now readily infer the required bound for the spectral gap of $\bar{P}^*$ (and hence also for $\bar{\mathsf{gap}}^\tau$).

Proof of Lemma 3.4.4. The following lemma of Chen [17] relates the contraction of the chain with its spectral gap:

Lemma 3.4.8 ([17])**.** *Let P be a transition kernel for a Markov chain on a metric space Ω. Suppose there exists a constant ι such that for each $x, y \in \Omega$, there is a coupling (X_1, Y_1) of*

$P(x, \cdot)$ and $P(y, \cdot)$ satisfying

$$\mathbf{E}_{x,y}(\mathrm{dist}(X_1, Y_1)) \leq \iota \,\mathrm{dist}(x, y) \ . \tag{3.4.8}$$

Then the spectral gap of P satisfies $\mathtt{gap} \geq 1 - \iota$.

By Lemma 4.2.8, the requirement (3.4.8) is satisfied with

$$\iota = \frac{b^\ell}{b^\ell + 1} + \frac{1}{b^\ell + 1} \cdot \frac{1 + (b-1)\ell}{b\kappa(1-\theta)(r-\ell)} \ ,$$

and hence

$$\begin{aligned}
\mathtt{gap}(\bar{P}^*) \geq 1 - \iota &= \frac{1}{b^\ell + 1}\left(1 - \frac{1 + (b-1)\ell}{b\kappa(1-\theta)(r-\ell)}\right) \\
&\geq \frac{1}{b^\ell + 1}\left(1 - \frac{\alpha}{\kappa(1-\theta)(1-2\alpha)}\right) \ ,
\end{aligned} \tag{3.4.9}$$

where in the last inequality we increased $1 + (b-1)\ell$ into $b\ell$ to simplify the final expression. This lower bound on $\mathtt{gap}(\bar{P}^*)$ translates via (3.4.4) into the desired lower bound on the spectral gap of the projection chain, $\bar{\mathtt{gap}}^\tau$. ∎

We are now ready to provide a lower bound on the spectral gap of the block dynamics, $\mathtt{gap}_{\mathcal{B}}^\tau$, and thereby conclude the proof of Theorem 3.4.1. By applying Theorem 3.2.6 to our decomposition of the block dynamics chain $P_{\mathcal{B}}^\tau$,

$$\mathtt{gap}_{\mathcal{B}}^\tau \geq \frac{\bar{\mathtt{gap}}}{3} \ \wedge \ \frac{\bar{\mathtt{gap}} \cdot \mathtt{gap}_{\min}}{3\gamma + \bar{\mathtt{gap}}} \ , \tag{3.4.10}$$

where

$$\mathtt{gap}_{\min} := \min_{\eta \in \{\pm 1\}^S} \mathtt{gap}_\eta^\tau \ , \qquad \gamma := \max_{\eta \in \{\pm 1\}^S} \max_{x \in \Omega_\eta} \sum_{y \in \Omega \backslash \Omega_\eta} P_{\mathcal{B}}^\tau(x, y) \ .$$

Lemma 3.4.2 gives that $\mathtt{gap}_{\min} \geq 1/(b^\ell + 1)$, and clearly, as the spins in S can only change if the distinguished block is updated, $\gamma \leq 1/(b^\ell + 1)$. Combining these two inequalities, we obtain that

$$\frac{\bar{\mathtt{gap}} \cdot \mathtt{gap}_{\min}}{3\gamma + \bar{\mathtt{gap}}} = \frac{\mathtt{gap}_{\min}}{1 + 3\gamma/\bar{\mathtt{gap}}} \geq \frac{1}{(b^\ell + 1) + 3/\bar{\mathtt{gap}}} \geq \frac{1}{4}\left(\frac{1}{b^\ell + 1} \wedge \bar{\mathtt{gap}}\right) \tag{3.4.11}$$

with room to spare. Together with (3.4.10), this implies that

$$\mathtt{gap}_{\mathcal{B}}^{\tau} \geq \frac{1}{4(b^{\ell}+1)} \ \wedge \ \frac{1}{4}\bar{\mathtt{gap}} \ ,$$

and Lemma 3.4.4 now gives that

$$\mathtt{gap}_{\mathcal{B}}^{\tau} \geq \frac{1}{4(b^{\ell}+1)}\left(1 - \frac{\alpha}{\kappa(1-\theta)(1-2\alpha)}\right) \ ,$$

as required. This concludes the proof of Theorem 3.4.1, and completes the proof of the upper bound on the mixing time. ∎

Remark 3. Throughout the proof of Theorem 6 we modified some of the constants (e.g., (3.4.9), (3.4.11) etc.) in order to simplify the final expressions obtained. By doing the calculations (slightly) more carefully, one can obtain an absolute constant of about 300 for the upper bound in Theorem 6.

3.5 Lower bounds on the mixing time and inverse-gap

In this section, we prove Theorem 8, which provides lower bounds on the inverse-gap and mixing time of the critical Ising model on a b-ary tree with free boundary. Throughout this section, let $b \geq 2$ be fixed, and set $\theta = \tanh \beta_c = \frac{1}{\sqrt{b}}$.

3.5.1 Lower bound on the inverse-gap

The required lower bound will be obtained by an application of the Dirichlet form (3.2.1), using a certain weighted sum of the spins as the corresponding test function.

Proof of Theorem 8, inequality (3.1.2). Let T be a b-ary tree, rooted at ρ, with h levels (and $n = \sum_{k=0}^{h} b^k$ vertices). We will show that

$$\mathtt{gap}^{-1} \geq \frac{b-1}{6b}nh^2 \ .$$

For simplicity, we use the abbreviation $d(v) := \mathrm{dist}(\rho, v)$, and define

$$g(\sigma) := \sum_{v \in T} \theta^{d(v)}\sigma(v) \ \ \text{for } \sigma \in \Omega \ .$$

By the Dirichlet form (3.2.1), and since $P(\sigma, \sigma') \leq \frac{1}{n}$ for any $\sigma, \sigma' \in \Omega$ in the discrete-time dynamics, we have that

$$\mathcal{E}(g) = \frac{1}{2} \sum_{\sigma, \sigma'} [g(\sigma) - g(\sigma')]^2 \mu(\sigma) P(\sigma, \sigma')$$

$$\leq \frac{1}{2} \max_{\sigma} \sum_{\sigma'} [g(\sigma) - g(\sigma')]^2 \mu(\sigma) P(\sigma, \sigma') \leq \frac{1}{2} \sum_{k=0}^{h} \frac{b^k}{n} (2\theta^k)^2 \leq \frac{2h}{n} .$$

On the other hand, the variance of g can be estimated as follows.

$$\mathrm{Var}_\mu \, g = \mathrm{Var}_\mu \left(\sum_{v \in T} \theta^{d(v)} \sigma(v) \right) = \sum_{u, w \in T} \theta^{d(u) + d(w)} \, \mathrm{Cov}_\mu(\sigma(u), \sigma(w))$$

$$= \sum_{u, v, w \in T} \theta^{d(u) + d(w)} \, \mathrm{Cov}_\mu(\sigma(u), \sigma(w)) \mathbf{1}_{\{u \wedge w = v\}} ,$$

where the notation $(u \wedge w)$ denotes their most immediate common ancestor (i.e., their common ancestor z with the largest $d(z)$). Notice that for each $v \in T$, the number of u, w that are of distance i, j from v respectively and have $v = u \wedge w$ is precisely $b^i \cdot (b-1)b^{j-1}$, since determining u immediately rules b^{j-1} candidates for w. Furthermore, by Claim 3.2.1 we have

$$\mathrm{Cov}_\mu(\sigma(u), \sigma(w)) = \theta^{d(u) + d(w) - 2d(v)} ,$$

and so

$$\mathrm{Var}_\mu \, g = \sum_{u, v, w \in T} \theta^{d(u) + d(w)} \theta^{d(u) + d(w) - 2d(v)} \mathbf{1}_{\{u \wedge w = v\}}$$

$$= \sum_{k=0}^{h} b^k \sum_{i=0}^{h-k} \sum_{j=0}^{h-k} b^i (b-1) b^{j-1} \theta^{2k+i+j} \theta^{i+j}$$

$$= \frac{b-1}{b} \sum_{k=0}^{h} (h-k)^2 = \frac{b-1}{6b} h(h+1)(2h+1) \geq \frac{b-1}{3b} h^3 .$$

Altogether, we can conclude that

$$\mathtt{gap} \leq \frac{\mathcal{E}(g)}{\mathrm{Var}_\mu \, g} = \frac{6b}{b-1} \cdot \frac{1}{nh^2} , \tag{3.5.1}$$

as required. $\blacksquare$
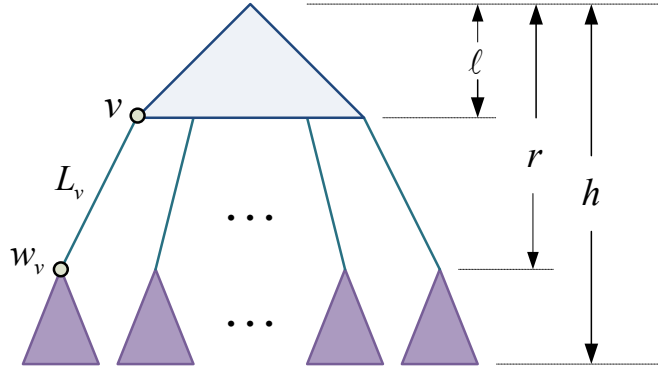
Figure 3.2: Speed-up dynamics for the Ising model on the tree.

3.5.2 Lower bound on the mixing-time

In order to obtain the required lower bound on the mixing time, we consider a "speed-up" version of the dynamics, namely a custom block-dynamics comprising a mixture of singletons and large subtrees. We will show that, even for this faster version of the dynamics, the mixing time has order at least $n \log^3 n$.

Let T be a b-ary tree with h levels (and $n = \sum_{k=0}^{h} b^k$ vertices). Consider two integers $1 \le \ell < r \le h$, to be specified later. For every $v \in H_\ell$, select one of its descendants in H_r arbitrarily, and denote it by w_v. Write $W = \{w_v : v \in H_\ell\}$ as the set of all such vertices. Further define

$$B_v := (T_v \setminus T_{w_v}) \cup \{w_v\} \qquad \text{(for each } v \in H_\ell) \ .$$

The *speed-up dynamics*, (X_t), is precisely the block-dynamics with respect to

$$\mathcal{B} = \{B_v : v \in H_\ell\} \cup \bigcup_{u \notin W} \{u\} \ .$$

In other words, the transition rule of the speed-up dynamics is the following:

(i) Select a vertex $u \in V(T)$ uniformly at random.

(ii) If $u \notin W$, update this site according to the usual rule of the Glauber dynamics.

(iii) Otherwise, update B_v given the rest of the spins, where $v \in H_\ell$ is the unique vertex with $u = w_v$.

The following theorem of [75] guarantees that, starting from all-plus configuration, the speed-up Glauber dynamics indeed mixes faster than the original one. In what follows, write $\mu \preceq \nu$ if μ stochastically dominates ν.
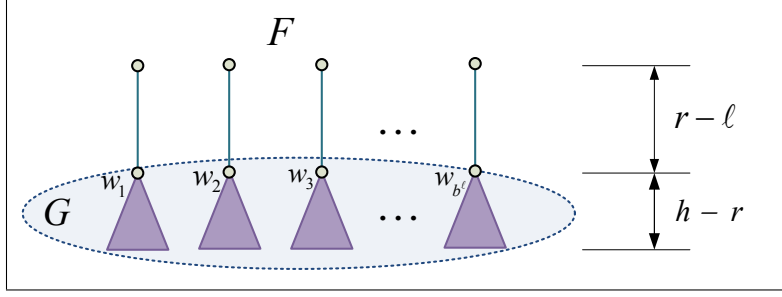
Figure 3.3: Speed-up dynamics on the forest F and the sub-forest G.

Theorem 3.5.1 ([75] and also see [74, Theorem 16.5]). *Let (Ω, S, V, π) be a monotone system and let μ be the distribution on Ω which results from successive updates at sites $v_1, \ldots, v_m$, beginning at the top configuration. Define ν similarly but with updates only at a subsequence $v_{i_1}, \ldots, v_{i_k}$. Then $\mu \preceq \nu$, and $\|\mu - \pi\|_{\mathrm{TV}} \leq \|\nu - \pi\|_{\mathrm{TV}}$. Moreover, this also holds if the sequence $v_1, \ldots, v_m$ and the subsequence $i_1, \ldots, i_k$ are chosen at random according to any prescribed distribution.*

To see that indeed the speed-up dynamics X_t is at least as fast as the usual dynamics, first note that any vertex $u \notin W$ is updated according to the original rule of the Glauber dynamics. Second, instead of updating the block B_v, we can simulate this operation by initially updating w_v (given its neighbors), and then performing sufficiently many single-site updates in B_v. This approximates the speed-up dynamics arbitrarily well, and comprises a superset of the single-site updates of the usual dynamics. The above theorem thus completes this argument.

It remains to estimate the mixing time of the speed-up dynamics X_t. To this end, define another set of blocks as follows: for every $v \in H_\ell$, let L_v denote the simple path between v and w_v (inclusive), define the forest

$$F := \bigcup_{v \in H_\ell} (L_v \cup T_{w_v}) \ ,$$

and put

$$\mathcal{B}_F := \{L_v : v \in H_\ell\} \cup \bigcup_{u \in F \setminus W} \{u\} \ .$$

We define Y_t, the speed-up dynamics on F, to be the block-dynamics with respect to $\mathcal{B}_F$ above. This should not be confused with running a dynamics on a subset of T with a boundary condition of the remaining vertices; rather than that, Y_t should be thought of as a dynamics on a separate graph F, which is endowed with a natural one-to-one mapping to

the vertices of T. Further note that, except for the singleton blocks in $\mathcal{B}$, every block $B_v \in \mathcal{B}$ in the block-dynamics X_t has a counterpart $L_v \subset B_v$ in Y_t.

The next lemma compares the continuous-time versions of X_t and Y_t (where each block is updated at rate 1), and shows that on a certain subset of the vertices, they typically remain the same for a substantial amount of time.

Lemma 3.5.2. *Let (X_t) and (Y_t) be the continuous-time speed-up dynamics on T and F respectively, as defined above. Let $G = \bigcup_{v \in H_\ell} T_{w_v}$ and define*

$$\tau = \inf_t \{ X_t(u) \neq Y_t(u) \text{ for some } u \in V(G) \} \ .$$

Then there exists a coupling of X_t and Y_t such that

$$\mathbf{P}(\tau > t) \geq \exp(-\theta^{r-\ell} b^\ell t) \ .$$

Proof. For two configurations $\sigma \in \{\pm 1\}^T$ and $\eta \in \{\pm 1\}^F$, denote their Hamming distance on F by

$$\text{dist}(\sigma, \eta) = \sum_{v \in F} \mathbf{1}_{\{\sigma(v) \neq \eta(v)\}} \ .$$

The coupling of X_t and Y_t up to time τ can be constructed as follows:

1. Whenever a singleton block $\{u\}$ with $u \in T \setminus F$ is being updated in X_t, the chain Y_t remains in place.

2. Otherwise, when a block B is updated in X_t, we update $B \cap F$ (the unique $B' \in \mathcal{B}_F$ with $B' \subset B$) in Y_t so as to minimize $\text{dist}(X_t, Y_t)$.

For any $w \in W$, define the stopping time

$$\tau_w = \inf\{ t : X_t(w) \neq Y_t(w) \} \ ,$$

and notice that in the above defined coupling we have $\tau = \min_{w \in W} \tau_w$, since W separates $G \setminus W$ from F.

Let $v \in H_\ell$ and $w = w_v \in W$, and suppose that block B_v is to be updated at time $t < \tau_w$ in X_t, and hence, as defined above, L_v is to be updated in Y_t. By definition, at this time these two blocks have the same boundary except for at v, where there is a boundary condition in T (the parent of v) and none in F (recall v is the root of one of the trees in F).

We now wish to give an upper bound on the probability that this update will result in $X_t(w) \neq Y_t(w)$. By the monotonicity of the Ising model, it suffices to give an upper bound for this event in the case where v has some parent z in F, and $X_t(z) \neq Y_t(z)$. In this case, we

can bound the probability that $X_t(w) \neq Y_t(w)$ (in the maximal coupling) by an expression of the form

$$\frac{1}{2}\left(\mu^{+,H}(\sigma(w)) - \mu^{-,H}(\sigma(w))\right)$$

as described in Lemma 3.4.7, where the external field H corresponds to the value of the spins in $T_w \setminus \{w\}$. Lemma 3.4.7 then allows us to omit the external field H at w, translating the problem into estimating the probability that a difference propagates from v to w. By Claim 3.2.1, we deduce that

$$\mathbf{P}\left(X_t(w) \neq Y_t(w)\right) \leq \theta^{r-\ell} ,$$

and therefore

$$\mathbf{P}\left(t < \tau_w\right) \geq \exp\left(-\theta^{r-\ell}t\right) .$$

Using the fact $|W| = b^\ell$, it follows that

$$\mathbf{P}(t < \tau) = \mathbf{P}\left(t < \min_{w \in W} \tau_w\right) \geq \exp(-\theta^{r-\ell}b^\ell t) ,$$

as required. ∎

With the above estimate on the probability that X_t and Y_t are equal on the subgraph G up to a certain time-point, we can now proceed to studying the projection of X_t on G via that of Y_t (being a product chain, Y_t is much simpler to analyze).

To be precise, let $\tilde{X}_t$ and $\tilde{Y}_t$ denote the respective projections of X_t and Y_t onto G, which as a reminder is the union of all trees T_{w_v}. Notice that $\tilde{Y}_t$ is precisely the continuous-time single-site Glauber dynamics on G, since the block update of L_v in F translates simply into the single-site update of w_v in G. On the other hand, $\tilde{X}_t$ is not even necessarily a Markov chain. We next prove a lower bound on the mixing time of the Markov chain $\tilde{Y}_t$.

Lemma 3.5.3. *Let $\tilde{H}_t$ be the transition kernel of $\tilde{Y}_t$, and let μ_G denote its corresponding stationary measure. Let $\mathtt{gap}'$ denote the spectral-gap of the continuous-time single-site dynamics on a b-ary tree of height $h - r$. Then*

$$\|\tilde{H}_t(\underline{1}, \cdot) - \mu_G\|_{\mathrm{TV}} > \frac{3}{5} \;\; \text{for any } t \leq \frac{\ell \log b - 2}{2\mathtt{gap}'} ,$$

where $\underline{1}$ denotes the all-plus configuration.

Proof. Let T' denote a b-ary tree of height $h - r$ and n' vertices. Let P' be the transition kernel of the corresponding discrete-time single-site Glauber dynamics on T', let H'_t be the transition kernel of the continuous-time version of this dynamics, and let μ' be their corresponding stationary measure.

By definition of G as a disjoint union of b^ℓ copies of T', clearly $\tilde{Y}_t$ is a product of b^ℓ copies of identical and independent component chains on T'. We can therefore reduce the analysis of $\tilde{Y}_t$ into that of H'_t, where the second eigenvalue of of its discrete-time counterpart P' plays a useful role.

The following lemma ensures that P' has an increasing eigenfunction corresponding to its second largest eigenvalue λ'.

Lemma 3.5.4 ([71, Lemma 3]). *The second eigenvalue of the discrete-time Glauber dynamics for the Ising model has an increasing eigenfunction.*

Since the eigenspace of λ' has an increasing eigenfunction, it also contains a monotone eigenfunction f such that $|f(\underline{1})| = \|f\|_\infty$. Therefore, the transition kernel of the continuous-time chain satisfies

$$
(H'_t f)(\underline{1}) = \left(\sum_{k=0}^\infty e^{-tn'} \frac{(tn')^k}{k!} (P')^k f \right)(\underline{1})
$$

$$
= e^{-tn'} \sum_{k=0}^\infty \frac{(tn'\lambda')^k}{k!} f(\underline{1}) = e^{-n'(1-\lambda')t} f(\underline{1}) . \tag{3.5.2}
$$

Since $\int f d\mu' = 0$, we have that

$$
|(H'_t f)(\underline{1})| = \left| \sum_y (H'_t(\underline{1}, y) f(y) - f(y)\mu'(y)) \right| \leq 2\|f\|_\infty \|H_t(\underline{1}, \cdot) - \mu'\|_{\text{TV}} .
$$

Plugging in (3.5.2) and using the fact that $|f(\underline{1})| = \|f\|_\infty$, it follows that

$$
\|H'_t(\underline{1}, \cdot) - \mu'\|_{\text{TV}} \geq \frac{1}{2} e^{-n'(1-\lambda')t} . \tag{3.5.3}
$$

In order to relate the product chain $\tilde{Y}_t$ to its component chain Y'_t, we will consider the Hellinger distance between certain distributions, defined next (for further details, see, e.g., [52]). First, define the *Hellinger integral* (also known as the *Hellinger affinity*) of two distribution μ and ν on Ω to be

$$
I_{\mathcal{H}}(\mu, \nu) := \sum_{x \in \Omega} \sqrt{\mu(x)\nu(x)} .
$$

The *Hellinger distance* is now defined as

$$
d_{\mathcal{H}}(\mu, \nu) := \sqrt{2 - 2I_{\mathcal{H}}(\mu, \nu)} .
$$

Clearly, for any two distributions μ and ν,

$$I_{\mathcal{H}}(\mu, \nu) = \sum_{x \in \Omega} \sqrt{\mu(x)\nu(x)} \geq \sum_{x \in \Omega} \mu(x) \wedge \nu(x) = 1 - \|\mu - \nu\|_{\mathrm{TV}} \ ,$$

and so $d_{\mathcal{H}}$ provides the following lower bound on the total variation distance:

$$\|\mu - \nu\|_{\mathrm{TV}} \geq 1 - I_{\mathcal{H}}(\mu, \nu) = \frac{1}{2} d_{\mathcal{H}}^2(\mu, \nu) \ . \tag{3.5.4}$$

Furthermore, the Hellinger distance also provides an upper bound on d_{TV}, as the next simple inequality (e.g., [36, Lemma 4.2 (i)]) shows:

$$\|\mu - \nu\|_{\mathrm{TV}} \leq d_{\mathcal{H}}(\mu, \nu) \ . \tag{3.5.5}$$

To justify this choice of a distance when working with product chains, notice that any two product measures $\mu = \prod_{i=1}^{n} \mu^{(i)}$ and $\nu = \prod_{i=1}^{n} \nu^{(i)}$ satisfy

$$I_{\mathcal{H}}(\mu, \nu) = \prod_{i=1}^{n} I_{\mathcal{H}}(\mu^{(i)}, \nu^{(i)}) \ . \tag{3.5.6}$$

Next, we consider the Hellinger integral of our component chains H_t'. Indeed, combining the definition of $d_{\mathcal{H}}$ with (3.5.5), we get that

$$I_{\mathcal{H}}(H_t'(\underline{1}, \cdot), \mu') \leq 1 - \frac{1}{2}\|H_t'(\underline{1}, \cdot) - \mu'\|_{\mathrm{TV}}^2 \leq 1 - \frac{1}{8}e^{-2n'(1-\lambda')t} \ ,$$

where the last inequality is by (3.5.3). Therefore, applying (3.5.6) to the product chain $\tilde{H}_t$ (the product of b^{ℓ} copies of H_t'), we can now deduce that

$$I_{\mathcal{H}}(\tilde{H}_t(\underline{1}, \cdot), \mu_G) \leq \left(1 - \frac{1}{8}e^{-2(1-\lambda')tn'}\right)^{b^{\ell}} \ .$$

At this point, (3.5.4) gives that

$$\|\tilde{H}_t(\underline{1}, \cdot) - \mu_G\|_{\mathrm{TV}} \geq 1 - \left(1 - \frac{e^{-2(1-\lambda')tn'}}{8}\right)^{b^{\ell}} \ .$$

Recall that by definition, $\mathtt{gap}'$ is the spectral-gap of H_t', the continuous-time version of P',

and so $\mathtt{gap}' = n'(1 - \lambda')$. Hence, if

$$t \leq \frac{\ell \log b - 2}{2\mathtt{gap}'}$$

then

$$\|\tilde{H}_t(\underline{1}, \cdot) - \mu_G\|_{\mathrm{TV}} \geq 1 - \exp\left(-e^2/8\right) > \frac{3}{5} ,$$

as required. ∎

The final ingredient required is the comparison between μ_G (the Gibbs distribution on G), and the projection of μ (the Gibbs distribution for T) onto the graph G. The following lemma provides an upper bound on the total-variation distance between these two measures.

Lemma 3.5.5. *Let μ and μ_G be the Gibbs distributions for T and G resp., and let $\tilde{\mu}$ denote the projection of μ onto G, that is:*

$$\tilde{\mu}(\eta) = \mu(\{\sigma \in \{\pm 1\}^T : \sigma_G = \eta\}) \qquad (\text{ for } \eta \in \{\pm 1\}^G) .$$

Then $\|\mu_G - \tilde{\mu}\|_{\mathrm{TV}} \leq b^{2\ell}\theta^{2(r-\ell)}$.

Proof. Recalling that G is a disjoint union of trees $\{T_w : w \in W\}$, clearly the configurations of these trees are independent according to μ_G. On the other hand, with respect to $\tilde{\mu}$, these configurations are correlated through their first (bottom-most) common ancestor. Further notice that, by definition, the distance between $w_i \neq w_j \in W$ in T is at least $2(r - \ell + 1)$, as they belong to subtrees of distinct vertices in H_ℓ.

To bound the effect of the above mentioned correlation, we construct a coupling between μ_G and $\tilde{\mu}$ iteratively on the trees $\{T_w : w \in W\}$, generating the corresponding configurations η and $\tilde{\eta}$, as follows. Order W arbitrarily as $W = \{w_1, \ldots, w_{b^\ell}\}$, and begin by coupling μ_G and $\tilde{\mu}$ on T_{w_1} via the identity coupling. Now, given a coupling on $\cup_{i<k}T_{w_i}$, we extend the coupling to T_{w_k} using a maximal coupling. Indeed, by essentially the same reasoning used for the coupling of the processes X_t and Y_t on G in Lemma 3.5.2, the probability that some already determined w_i (for $i < k$) would affect w_k is at most $\theta^{2(r-\ell+1)}$. Summing these probabilities, we have that

$$\mathbf{P}\left(\eta_{T_{w_k}} \neq \tilde{\eta}_{T_{w_k}}\right) = \mathbf{P}\left(\eta(w_k) \neq \tilde{\eta}(w_k)\right) \leq (k-1)\theta^{2(r-\ell+1)} .$$

Altogether, taking another union bound over all $k \in [b^\ell]$, we conclude that

$$\|\mu_G - \tilde{\mu}\|_{\mathrm{TV}} \leq \mathbf{P}(\eta \neq \tilde{\eta}) \leq b^{2\ell}\theta^{2(r-\ell)} ,$$

completing the proof. ∎

We are now ready to prove the required lower bound on t_{MIX}.

Proof of Theorem 8, inequality (3.1.3). As we have argued above (see Theorem 4.2.4 and the explanation thereafter), it suffices to establish a lower bound on the mixing time of the speed-up dynamics X_t on T. By considering the projection of this chain onto G, we have that

$$\|\mathbf{P}_{\underline{1}}(X_t \in \cdot) - \mu\|_{\mathrm{TV}} \geq \|\mathbf{P}_{\underline{1}}(\tilde{X}_t \in \cdot) - \tilde{\mu}\|_{\mathrm{TV}} \ ,$$

and recalling the definition of τ as $\inf_t\{(X_t)_G \neq (Y_t)_G\}$,

$$\|\mathbf{P}_{\underline{1}}(\tilde{X}_t \in \cdot) - \tilde{\mu}\|_{\mathrm{TV}} \geq \|\mathbf{P}_{\underline{1}}(\tilde{Y}_t \in \cdot) - \tilde{\mu}\|_{\mathrm{TV}} - \mathbf{P}(\tau \leq t)$$
$$\geq \|\mathbf{P}_{\underline{1}}(\tilde{Y}_t \in \cdot) - \mu_G\|_{\mathrm{TV}} - \mathbf{P}(\tau \leq t) - \|\mu_G - \tilde{\mu}\|_{\mathrm{TV}} \ .$$

Let `gap` and `gap`$'$ denote the spectral-gaps of the continuous-time single-site dynamics on a b-ary tree with h levels and $h - r$ levels respectively (and free boundary condition), and choose t such that

$$t \leq \frac{\ell \log b - 2}{2 \mathtt{gap}'} \ . \tag{3.5.7}$$

Applying Lemmas 3.5.2, 3.5.3 and 3.5.5, we obtain that

$$\|\mathbf{P}_{\underline{1}}(X_t \in \cdot) - \mu\|_{\mathrm{TV}} \geq \frac{3}{5} - \left(1 - \exp(-\theta^{r-\ell} b^\ell t)\right) - b^{2\ell}\theta^{2(r-\ell)} \ . \tag{3.5.8}$$

Now, selecting

$$\ell = \frac{h}{5} \quad \text{and} \quad r = \frac{4h}{5} \ ,$$

and recalling that $b\theta^2 = 1$, we have that the last two terms in (3.5.8) both tend to 0 as $h \to \infty$, and so

$$\|P_{\underline{1}}(X_t \in \cdot) - \mu\|_{\mathrm{TV}} \geq \frac{3}{5} - o(1) \ .$$

In particular, for a sufficiently large h, this distance is at least $1/e$, hence by definition the continuous-time dynamics satisfies $t_{\mathrm{MIX}} \geq t$. We can can now plug in our estimates for `gap`$'$ to obtain the required lower bounds on t_{MIX}.

First, recall that by (3.5.1),

$$\mathtt{gap}' \leq \frac{6b}{b-1} \cdot \frac{1}{(h-r)^2} \ ,$$

and so the following choice of t satisfies (3.5.7):

$$t := \frac{(b-1)}{12b}(h-r)^2(\ell \log b - 2) \ .$$

It follows that the mixing-time of the continuous-time dynamics satisfies

$$t_{\text{MIX}} \geq t \geq \left(\frac{(b-1)\log b}{1500\,b} + o(1)\right)h^3 \ ,$$

and the natural translation of this lower bound into the discrete-time version of the dynamics yields the lower bound in (3.1.3).

Second, let $g(h)$ be the continuous-time inverse-gap of the dynamics on the b-ary tree of height h with free-boundary condition, and recall that by Theorem 6, we have that g is polynomial in h. In particular,

$$g(h) \leq Cg(h/5) \quad \text{for some fixed } C > 0 \text{ and all } h.$$

Since by definition $(\mathtt{gap}')^{-1} = g(h-r) = g(h/5)$ and $\mathtt{gap}^{-1} = g(h)$, we can choose t to be the right-hand-side of (3.5.7) and obtain that for any large h

$$t_{\text{MIX}} \geq t \geq C'\mathtt{gap}^{-1}h \quad \text{for some } C' > 0 \text{ fixed.}$$

Clearly, this statement also holds when both t_{MIX} and $\mathtt{gap}$ correspond to the discrete-time version of the dynamics, completing the proof. ∎

3.6 Phase transition to polynomial mixing

This section contains the proof of Theorem 7, which addresses the near critical Ising model on the tree, and namely, the transition of its (continuous-time) inverse-gap and mixing-time from polynomial to exponential in the tree-height. Theorem 7 will follow directly from the next theorem:

Theorem 3.6.1. *Fix $b \geq 2$, let $\varepsilon = \varepsilon(h)$ satisfy $0 < \varepsilon < \varepsilon_0$ for a suitably small constant ε_0, and let $\beta = \operatorname{arctanh}\left(\sqrt{(1+\varepsilon)/b}\right)$. The following holds for the continuous-time Glauber dynamics for the Ising model on the b-ary tree with h levels at the inverse-temperature β:*

(i) For some $c_1 > 0$ fixed, the dynamics with free boundary satisfies

$$\mathtt{gap}^{-1} \geq c_1\left((1/\varepsilon) \wedge h\right)^2(1+\varepsilon)^h \ . \tag{3.6.1}$$

(ii) For some absolute constant $c_2 > 0$ and any boundary condition τ

$$\mathtt{gap}^{-1} \leq t_{\mathrm{MIX}} \leq \mathrm{e}^{c_2(\varepsilon h + \log h)} \ . \tag{3.6.2}$$

Throughout this section, let $b \geq 2$ be some fixed integer, and let T be a b-ary tree with height h and n vertices. Define $\theta = \sqrt{(1+\varepsilon)/b}$, where $\varepsilon = \varepsilon(n)$ satisfies $0 < \varepsilon \leq \varepsilon_0$ (for some suitably small constant $\varepsilon_0 < \frac{1}{8}$ to be later specified), and as usual write $\beta = \mathrm{arctanh}(\theta)$.

Proof of Theorem 3.6.1. The proof follows the same arguments of the proof of Theorems 6 and 8. Namely, the upper bound uses an inductive step using a similar block dynamics, and the decomposition of this chain to establish a bound on its gap (as in Section 3.4) via the spatial mixing properties of the Ising model on the tree (studied in Section 3.3). The lower bound will again follow from the Dirichlet form, using a testing function analogous to the one used in Section 3.5. As most of the arguments carry to the new regime of β in a straightforward manner, we will only specify the main adjustments one needs to make in order to extend Theorems 6 and 8 to obtain Theorem 3.6.1.

Upper bound on the inverse-gap

Let $\frac{1}{100} < \kappa < 1$ be the universal constant that was introduced in Lemma 3.3.7 (and appears in Proposition 3.3.1 and Theorem 3.3.2), and define

$$\varepsilon_0 := \frac{\kappa}{20} \leq \frac{1}{8} \ .$$

As $b \geq 2$ and $\varepsilon < \varepsilon_0 \leq \frac{1}{8}$, we have that $\theta \leq \frac{3}{4}$, hence Proposition 3.3.1 and Theorem 3.3.2 both hold in this supercritical setting. It therefore remains to extend the arguments in Section 3.4 (that use Proposition 3.3.1 as one of the ingredients in the proof of the upper bound on $\mathtt{gap}^{-1}$) to this new regime of β.

Begin by defining the same block dynamics as in (3.4.2), only with respect to the following choice of ℓ and r (replacing their definition (3.4.1)):

$$\alpha := \varepsilon_0 = \kappa/20 \ , \tag{3.6.3}$$

$$\ell := \alpha \left[(1/\varepsilon) \wedge h \right] \ , \ r := h - \ell \ . \tag{3.6.4}$$

Following the same notations of Section 3.4, we now need to revisit the arguments of Lemma 4.2.8, and extend them to the new value of $\theta = \tanh \beta = \sqrt{(1+\varepsilon)/b}$. This comprises the following two elements:

1. Bounding the L^2-capacity $\mathrm{cap}_2(B(\varrho, r - \ell))$.

2. Estimating the probability that a difference in one spin would propagate to other spins, when coupling two instances of the chain $\bar{P}^*$.

Recalling the Nash-Williams Criterion (Lemma 3.4.6) and its application in inequality (3.4.6), the effective resistance between ϱ and $\partial B(\varrho, r - \ell)$ is at least

$$\sum_{k=1}^{r-\ell+1} \left(b^k \theta^{2k}\right)^{-1} = \sum_{k=1}^{r-\ell+1} (1+\varepsilon)^{-k} = \frac{1}{\varepsilon}\left(1 - (1+\varepsilon)^{-(r-\ell+1)}\right) \ ,$$

which implies that

$$\mathrm{cap}_2\left(B(\varrho, r - \ell + 1)\right) \leq \frac{\varepsilon}{1 - (1+\varepsilon)^{-(r-\ell)}} \ . \tag{3.6.5}$$

Now, if $\varepsilon \geq 1/h$, we have

$$1 - (1+\varepsilon)^{-(r-\ell)} = 1 - (1+\varepsilon)^{-(h-2\alpha/\varepsilon)} \geq 1 - (1+\varepsilon)^{-(1-2\alpha)/\varepsilon}$$
$$\geq 1 - e^{-(1-2\alpha)} \geq \frac{1 - 2\alpha}{2} \ ,$$

where the last inequality uses the fact that $\exp(-x) \leq 1 - x + \frac{x^2}{2}$ and that $\alpha > 0$. Similarly, if $\varepsilon < 1/h$ then

$$1 - (1+\varepsilon)^{-(r-\ell)} = 1 - (1+\varepsilon)^{-h(1-2\alpha)} \geq 1 - e^{-\varepsilon h(1-2\alpha)}$$
$$\geq \varepsilon h(1 - 2\alpha) - \frac{(\varepsilon h(1 - 2\alpha))^2}{2} \geq \varepsilon h \frac{1 - 2\alpha}{2} \ ,$$

where in the last inequality we plugged in the fact that $\varepsilon h < 1$. Combining the last two equations with (3.6.5), we deduce that

$$\mathrm{cap}_2\left(B(\varrho, r - \ell + 1)\right) \leq \frac{2\left(\varepsilon \vee (1/h)\right)}{1 - 2\alpha} \ .$$

Using (3.4.5), it then follows that

$$\mathbf{E}_{\varphi,\psi}\left(\bar{X}_1^*(\varrho) - \bar{Y}_1^*(\varrho) \,\middle|\, E_2\right) \leq \frac{2\left(\varepsilon \vee (1/h)\right)}{\kappa(1 - \theta)(1 - 2\alpha)} \ .$$

By repeating the next arguments of Lemma 4.2.8 (without any additional essential changes),

we obtain that under the monotone coupling,

$$
\begin{aligned}
\mathbf{E}_{\varphi,\psi}\Big(\mathrm{dist}(\bar{X}_1^*,\bar{Y}_1^*) \,\big|\, E_2\Big) &\leq \frac{2\,(\varepsilon \,\vee\, (1/h))}{\kappa(1-\theta)(1-2\alpha)}\Big(1 + \sum_{k=1}^{\ell-1}\frac{b-1}{b}b^k\theta^{2k}\Big)\\
&= \frac{2\,(\varepsilon \,\vee\, (1/h))}{\kappa(1-\theta)(1-2\alpha)}\Big(\frac{1}{b} + \frac{b-1}{b}\frac{(1+\varepsilon)^{\alpha[(1/\varepsilon)\wedge h]}-1}{\varepsilon}\Big)\\
&\leq \frac{2\,(\varepsilon \,\vee\, (1/h))}{\kappa(1-\theta)(1-2\alpha)}\frac{(1+\varepsilon)^{\alpha[(1/\varepsilon)\wedge h]}-1}{\varepsilon} \leq \frac{2\,(\varepsilon \,\vee\, (1/h))}{\kappa(1-\theta)(1-2\alpha)}\frac{e^{\alpha[1\wedge \varepsilon h]}-1}{\varepsilon}\\
&\leq \frac{2\,(\varepsilon \,\vee\, (1/h))}{\kappa(1-\theta)(1-2\alpha)}\frac{2\alpha[1\wedge \varepsilon h]}{\varepsilon} = \frac{4\alpha}{\kappa(1-\theta)(1-2\alpha)}\ ,
\end{aligned}
$$

where in the last line we used the fact that $e^x - 1 < 2x$ for all $0 \leq x \leq 1$. Again defining $g_h = n_h \min_\tau \mathtt{gap}_h^\tau$, we note that all the remaining arguments in Section 3.4 apply in our case without requiring any modifications, hence the following recursion holds for g_h:

$$
g_h \geq c(\alpha)g_r = c(\alpha)g_{h-\alpha[(1/\varepsilon)\wedge h]}\ , \tag{3.6.6}
$$

where

$$
c(\alpha) := \frac{1}{8}\Big(1 - \frac{4\alpha}{\kappa(1-\theta)(1-2\alpha)}\Big)\ .
$$

Recalling the definition (3.6.3) of α, since $\theta \leq \frac{3}{4}$ and $\kappa < 1$ we have that

$$
\frac{4\alpha}{\kappa(1-\theta)(1-2\alpha)} = \frac{2}{(1-\theta)(10-\kappa)} < \frac{8}{9}\ ,
$$

and so $c(\alpha) > 0$. We now apply the next recursion over g_{h_k}:

$$
h_0 = h\ ,\ \ h_{k+1} = \begin{cases} h_k - (\alpha/\varepsilon) & \text{if } h_k \geq (1/\varepsilon)\ ,\\ (1-\alpha)h_k & \text{if } h_k \leq (1/\varepsilon)\ . \end{cases}
$$

Notice that by our definition (3.6.3), we have $\varepsilon < \varepsilon_0 = \alpha$. With this in mind, definition (3.6.4) now implies that for any $h > 1/\alpha$ we have $\ell, r \geq 1$. Thus, letting $K = \min\{k : h_k \leq 1/\alpha\}$, we can conclude from (3.6.6) that

$$
\begin{aligned}
&g_{h_k} \geq c(\alpha)g_{h_{k+1}}\ \text{for all } k < K,\ \text{and hence}\\
&g_h \geq (c(\alpha))^K g_{h_K}\ .
\end{aligned}
$$

By the definitions of h_k and K, clearly

$$K \leq \frac{\varepsilon}{\alpha} h + \log_{1/(1-\alpha)} \left(h \wedge (1/\varepsilon) \right) = O(\varepsilon h + \log h) \ .$$

Since $h_K \leq 1/\alpha$, clearly $g_{h_K} > c'$ for some constant $c' = c'(\alpha) > 0$, giving

$$g_h \geq c' \left(c(\alpha) \right)^K \geq \mathrm{e}^{-M(\varepsilon h + \log h)}$$

for some constant $M = M(\alpha) > 0$ and any sufficiently large n. By definition of g_h, this provides an upper bound on $\mathtt{gap}^{-1}$, and as $t_{\mathrm{MIX}} = O\left(\mathtt{gap}^{-1} \log^2 n \right)$ (see Corollary 3.2.4 in Section 5.1), we obtain the upper bound on t_{MIX} that appears in (3.6.2).

Lower bound on the inverse-gap

We now turn to establishing a lower bound on the inverse-gap. Define the test function g to be the same one given in Subsection 3.5.1:

$$g(\sigma) = \sum_{v \in T} \theta^{\mathrm{dist}(\rho, v)} \sigma(v) \ .$$

By the same calculations as in the proof of Theorem 8 (Subsection 3.5.1), we have that

$$\mathcal{E}(g) \leq \frac{1}{2} \sum_{k=0}^{h} \frac{b^k}{n} (2\theta^k)^2 = \frac{2}{n} \sum_{k=0}^{h} (1+\varepsilon)^k = \frac{2}{n} \frac{(1+\varepsilon)^{h+1} - 1}{\varepsilon} \ , \tag{3.6.7}$$

whereas

$$\begin{aligned}
\mathrm{Var}_\mu(g) &= \frac{b-1}{b} \sum_{k=0}^{h} b^k \theta^{2k} \left(\sum_{i=0}^{h-k} b^i \theta^{2i} \right)^2 = \frac{b-1}{b} \sum_{k=0}^{h} (1+\varepsilon)^k \left(\sum_{i=0}^{h-k} (1+\varepsilon)^i \right)^2 \\
&= \frac{b-1}{b} \sum_{k=0}^{h} (1+\varepsilon)^k \left(\frac{(1+\varepsilon)^{h-k+1} - 1}{\varepsilon} \right)^2 \\
&= \frac{b-1}{b\,\varepsilon^2} \sum_{k=0}^{h} \left((1+\varepsilon)^{2h-k+2} - 2(1+\varepsilon)^{h+1} + (1+\varepsilon)^k \right) \\
&= \frac{b-1}{b\,\varepsilon^3} \left((1+\varepsilon)^{2h+3} - (2h+3)\,\varepsilon(1+\varepsilon)^{h+1} - 1 \right) \ . \tag{3.6.8}
\end{aligned}$$

When $\varepsilon \geq 8/h$ we have

$$\frac{1}{2}(1+\varepsilon)^{h+2} - (2h+3)\varepsilon \geq \frac{\varepsilon}{2}(h+2) + \frac{\varepsilon^2}{2}\binom{h+2}{2} - (2h+3)\varepsilon$$

$$\geq (h+2)\varepsilon\left(\frac{1}{2} + \varepsilon\frac{h+1}{4} - 2\right) \geq 4 ,$$

and therefore in this case (3.6.8) gives

$$\mathrm{Var}_\mu(g) \geq \frac{b-1}{2b}\frac{(1+\varepsilon)^{2h+3}}{\varepsilon^3} . \tag{3.6.9}$$

Combining (3.6.7) and (3.6.9), the Dirichlet form (3.2.1) now gives that

$$\mathtt{gap} \leq \frac{4b}{b-1}\frac{\varepsilon^2}{n(1+\varepsilon)^h} \quad \text{for } \varepsilon \geq 8/h . \tag{3.6.10}$$

On the other hand, when $0 \leq \varepsilon < 8/h$ we still have $b\theta^2 \geq 1$ and hence

$$\mathrm{Var}_\mu(g) = \frac{b-1}{b}\sum_{k=0}^{h} b^k\theta^{2k}\left(\sum_{i=0}^{h-k} b^i\theta^{2i}\right)^2 \geq \frac{b-1}{b}\sum_{k=0}^{h}(h-k)^2 \geq \frac{b-1}{3b}h^3.$$

In addition, using the fact that the expression $[(1+\varepsilon)^{h+1} - 1]/\varepsilon$ in (3.6.7) is monotone increasing in ε, in this case we have

$$\mathcal{E}(g) \leq \frac{2}{n}\frac{(1+(8/h))^{h+1} - 1}{8/h} \leq \mathrm{e}^7 h/n ,$$

where the last inequality holds for any $h \geq 20$. Altogether, the Dirichlet form (3.2.1) yields (for such values of h)

$$\mathtt{gap} \leq \frac{3\mathrm{e}^7 b}{b-1}\frac{1}{nh^2} \quad \text{for } 0 < \varepsilon \leq 8/h . \tag{3.6.11}$$

Combining (3.6.10) and (3.6.11), we conclude that

$$\mathtt{gap} \leq \frac{3\mathrm{e}^{15} b}{b-1}\left[n(1+\varepsilon)^h\left((1/\varepsilon) \wedge h\right)^2\right]^{-1} ,$$

where we used the fact that $(1+\varepsilon)^h \leq \mathrm{e}$. This gives the lower bound on $\mathtt{gap}^{-1}$ that appears in (3.6.1), completing the proof of Theorem 3.6.1. ∎

3.7 Concluding remarks and open problems

- We have established that in the continuous-time Glauber dynamics for the critical Ising model on a regular tree with arbitrary boundary condition, both the inverse-gap and the mixing-time are polynomial in the tree-height h. This completes the picture for the phase-transition of the inverse-gap (bounded at high temperatures, polynomial at criticality and exponential at low temperatures), as conjectured by the physicists for lattices. Moreover, this provides the first proof of this phenomenon for any underlying geometry other than the complete graph.

- In addition, we studied the near-critical behavior of the inverse-gap and mixing-time. Our results yield the critical exponent of $\beta - \beta_c$, as well as pinpoint the threshold at which these parameters cease to be polynomial in the height.

- For further study, it would now be interesting to determine the precise power of h in the order of each the parameters $\mathtt{gap}^{-1}$ and t_{MIX} at the critical temperature. In the free-boundary case, our lower bounds for these parameters in Theorem 8 provide candidates for these exponents:

  Question 3.7.1. Fix $b \geq 2$ and let $\beta_c = \mathrm{arctanh}(1/\sqrt{b})$ be the critical inverse-temperature for the Ising model on a b-ary tree of height h. Does the corresponding continuous-time Glauber dynamics with free boundary condition satisfy $\mathtt{gap}^{-1} \asymp h^2$ and $t_{\mathrm{MIX}} \asymp h^3$?

- Both at critical and at near-critical temperatures, our upper bounds for the inverse-gap and mixing-time under an arbitrary boundary condition matched the behavior in the free-boundary case. This suggests that a boundary condition can only accelerate the mixing of the dynamics, and is further supported by the behavior of the model under the all-plus boundary, as established in [66]. We therefore conjecture the following monotonicity of $\mathtt{gap}^{-1}$ and t_{MIX} with respect to the boundary condition:

  Conjecture 1. *Fix $b \geq 2$ and $\beta > 0$, and consider the Ising model on a b-ary tree with parameter β. Denote by $\mathtt{gap}$ and t_{MIX} the spectral-gap and mixing time for the Glauber dynamics with free boundary, and by $\mathtt{gap}^\tau$ and t_{MIX}^τ those with boundary condition τ. Then*

  $$\mathtt{gap} \leq \mathtt{gap}^\tau \quad and \quad t_{\mathrm{MIX}} \geq t_{\mathrm{MIX}}^\tau \quad for\ any\ \tau.$$

- A related statement was proved in [65] for two-dimensional lattices at low temperature: It was shown that, in that setting, the spectral-gap under the all-plus boundary condition is substantially larger than the spectral-gap under the free boundary condition. In light of this, it would be interesting to verify whether the monotonicity property, described in Conjecture 1, holds for the Ising model on an arbitrary finite graph.

Chapter 4

General lower bound on the mixing for Ising model

4.1 Introduction

Consider a finite graph $G = (V, E)$ and a finite alphabet Q. A general *spin system* on G is a probability measure μ on Q^V; well studied examples in computer science and statistical physics include the uniform measure on proper colorings and the Ising model. Glauber (heat-bath) dynamics are often used to sample from μ (see, e.g., [54, 64, 80]). In discrete-time Glauber dynamics, at each step a vertex v is chosen uniformly at random and the label at v is replaced by a new label chosen from the μ-conditional distribution given the labels on the other vertices. This Markov chain has stationary distribution μ, and the key quantity to analyze is the mixing time t_{MIX}, at which the distribution of the chain is close in total variation to μ (precise definitions are given below).

If $|V| = n$, it takes $(1 + o(1))n \log n$ steps to update all vertices (coupon collecting), and it is natural to guess that this is a lower bound for the mixing time. However, for the Ising model at infinite temperature or equivalently, for the 2-colorings of the graph $(V, \emptyset)$, the mixing time of Glauber dynamics is asymptotic to $n \log n/2$, since these models reduce to the lazy random walk on the hypercube, first analyzed in [2]. Thus mixing can occur before all sites are updated, so the coupon collecting argument does not suffice to obtain a lower bound for the mixing time. The first general bound of the right order was obtained by Hayes and Sinclair [42], who showed that the mixing time for Glauber dynamics is at least $n \log n/f(\Delta)$, where Δ is the maximum degree and $f(\Delta) = \Theta(\Delta \log^2 \Delta)$. Their result applies for quite general spin systems, and they gave examples of spin systems μ where some dependence on Δ is necessary. After the work of [42], it remained unclear whether a uniform

lower bound of order $n \log n$, *that does not depend on* Δ, holds for the most extensively studied spin systems, such as proper colorings and the Ising model.

In this chapter, we focus on the ferromagnetic Ising model, and obtain a lower bound of $(1/4 + o(1))n \log n$ on any graph with general (non-negative) interaction strengths.

Recall the definition of the Ising measure as in (1.1.1). Throughout this chapter, we take $\beta = 1$ and $H \equiv 0$ unless otherwise specified. We now state the main result of this chapter.

Theorem 9. *Consider the Ising model (1.1.1) on the graph G with interaction matrix J, and let $t^+_{\mathrm{MIX}}(G, J)$ denote the mixing time of the corresponding Glauber dynamics, started from the all-plus configuration. Then*

$$\inf_{G,J} t^+_{\mathrm{MIX}}(G, J) \geq (1/4 + o(1))n \log n \,,$$

where the infimum is over all n-vertex graphs G and all nonnegative interaction matrices J.

Remark. Theorem 9 is sharp up to a factor of 2. We conjecture that $(1/4 + o(1))$ in the theorem could be replaced by $(1/2 + o(1))$, i.e., the mixing time is minimized (at least asymptotically) by taking $J \equiv 0$.

Hayes and Sinclair [42] constructed spin systems where the mixing time of the Glauber dynamics has an upper bound $O(n \log n / \log \Delta)$. This, in turn, implies that in order to establish a lower bound of order $n \log n$ for the Ising model on a general graph, we have to employ some specific properties of the model. In our proof of Theorem 9, given in the next section, we use the GHS inequality [40] (see also [51] and [32]) and a recent censoring inequality [75] due to Peter Winkler and the second author.

4.2   Proof of Theorem 9

The intuition for the proof is the following: In the case of strong interactions, the spins are highly correlated and the mixing should be quite slow; In the case of weak interaction strengths, the spins should be weakly dependent and close to the case of the graph with no edges, therefore one may extend the arguments for the lazy walk on the hypercube.

We separate the two cases by considering the *spectral gap*. Recall that the spectral gap of a reversible discrete-time Markov chain, denoted by $\mathtt{gap}$, is $1 - \lambda$, where λ is the second largest eigenvalue of the transition kernel. The following simple lemma gives a lower bound on t^+_{MIX} in terms of the spectral gap.

Lemma 4.2.1. *The Glauber dynamics for the ferromagnetic Ising model (1.1.1) satisfies* $t^+_{\mathrm{MIX}} \geq \log 2 \cdot (\mathtt{gap}^{-1} - 1)$.

Proof. It is well known that $t_{\text{MIX}} \geq \log 2 \cdot (\text{gap}^{-1} - 1)$ (see, e.g., Theorem 12.4 in [54]). Actually, it is shown in the proof of [54, Theorem 12.4] that $t_{\text{MIX}}^x \geq \log 2 \cdot (\text{gap}^{-1} - 1)$ for any state x satisfying $f(x) = \|f\|_\infty$, where f is an eigenfunction corresponding to the second largest eigenvalue. Since the second eigenvalue of the Glauber dynamics for the ferromagnetic Ising model has an increasing eigenfunction f (see [71, Lemma 3]), we infer that either $\|f\|_\infty = f(+)$ or $\|f\|_\infty = f(-)$. By symmetry of the all-plus and the all-minus configurations in the Ising model (1.1.1), we have $t_{\text{MIX}}^+ = t_{\text{MIX}}^-$, and this concludes the proof. ∎

Lemma 4.2.1 implies that Theorem 9 holds if $\text{gap}^{-1} \geq n \log n$. It remains to consider the case $\text{gap}^{-1} \leq n \log n$.

Lemma 4.2.2. *Suppose that the Glauber dynamics for the Ising model on a graph $G = (V, E)$ with n vertices satisfies $\text{gap}^{-1} \leq n \log n$. Then there exists a subset $F \subset V$ of size $\lfloor \sqrt{n}/\log n \rfloor$ such that*

$$\sum_{u,v \in F, u \neq v} \text{Cov}_\mu(\sigma(u), \sigma(v)) \leq \frac{2}{\log n} .$$

Proof. We first establish an upper bound on the variance of the sum of spins $S = S(\sigma) = \sum_{v \in V} \sigma(v)$. The variational principle for the spectral gap of a reversible Markov chain with stationary measure π gives (see, e.g., [3, Chapter 3] or [54, Lemma 13.12]:

$$\text{gap} = \inf_f \frac{\mathcal{E}(f)}{\text{Var}_\pi(f)} ,$$

where $\mathcal{E}(f)$ is the Dirichlet form defined by

$$\mathcal{E}(f) = \langle (I - P)f, f \rangle_\pi = \tfrac{1}{2} \sum_{x,y \in \Omega} [f(x) - f(y)]^2 \pi(x) P(x,y) .$$

Applying the variational principle with the test function S, we deduce that

$$\text{gap} \leq \frac{\mathcal{E}(S)}{\text{Var}_\mu(S)} .$$

Since the Glauber dynamics updates a single spin at each step, $\mathcal{E}(S) \leq 2$, whence

$$\text{Var}_\mu(S) \leq \mathcal{E}(S)\text{gap}^{-1} \leq 2\,n \log n . \tag{4.2.1}$$

The covariance of the spins for the ferromagnetic Ising model is non-negative by the FKG inequality (see, e.g., [39]). Applying Claim 4.2.3 below with $k = \lfloor \frac{\sqrt{n}}{\log n} \rfloor$ to the covariance matrix of σ concludes the proof of the lemma. ∎

Claim 4.2.3. *Let A be an $n \times n$ matrix with non-negative entries. Then for any $k \leq n$ there exists $F \subset \{1, \ldots, n\}$ such that $|F| = k$ and*

$$\sum_{i,j \in F} A_{i,j} \mathbf{1}_{\{i \neq j\}} \leq \frac{k^2}{n^2} \sum_{i \neq j} A_{i,j} .$$

Proof. Let R be a uniform random subset of $\{1, \ldots, n\}$ with $|R| = k$. Then,

$$\mathbf{E}\Big[\sum_{i,j \in R} A_{i,j} \mathbf{1}_{\{i \neq j\}}\Big] = \sum_{1 \leq i,j \leq n} A_{i,j} \mathbf{1}_{\{i \neq j\}} \mathbf{P}(i, j \in R)$$

$$= \frac{k(k-1)}{n(n-1)} \sum_{1 \leq i,j \leq n} A_{i,j} \mathbf{1}_{\{i \neq j\}} \leq \frac{k^2}{n^2} \sum_{i \neq j} A_{i,j} .$$

Existence of the desired subset F follows immediately. ∎

We now consider a version of accelerated dynamics (X_t) with respect to the subset F as in Lemma 4.2.2. The accelerated dynamics selects a vertex $v \in V$ uniformly at random at each time and updates in the following way:

- If $v \notin F$, we update $\sigma(v)$ as in the usual Glauber dynamics.

- If $v \in F$, we update the spins on $\{v\} \cup F^c$ all together as a block, according to the conditional Gibbs measure given the spins on $F \setminus \{v\}$.

The next censoring inequality for monotone systems of [75] guarantees that, starting from the all-plus configuration, the accelerated dynamics indeed mixes faster than the original one. A *monotone system* is a Markov chain on a partially ordered set with the property that for any pair of states $x \leq y$ there exist random variables $X_1 \leq Y_1$ such that for every state z

$$\mathbf{P}(X_1 = z) = p(x, z), \quad \mathbf{P}(Y_1 = z) = p(y, z) .$$

In what follows, write $\mu \preceq \nu$ if ν stochastically dominates μ.

Theorem 4.2.4 ([75] and also see [74, Theorem 16.5])**.** *Let (Ω, S, V, π) be a monotone system and let μ be the distribution on Ω which results from successive updates at sites $v_1, \ldots, v_m$, beginning at the top configuration. Define ν similarly but with updates only at a subsequence $v_{i_1}, \ldots, v_{i_k}$. Then $\mu \preceq \nu$, and $\|\mu - \pi\|_{\mathrm{TV}} \leq \|\nu - \pi\|_{\mathrm{TV}}$. Moreover, this also holds if the sequence $v_1, \ldots, v_m$ and the subsequence $i_1, \ldots, i_k$ are chosen at random according to any prescribed distribution.*

In order to see how the above theorem indeed implies that the accelerated dynamics (X_t) mixes at least as fast as the usual dynamics, first note that any vertex $u \notin F$ is updated according to the original rule of the Glauber dynamics. Second, for $u \in F$, instead of updating the block $\{u\} \cup F^c$, we can simulate this procedure by performing sufficiently many single-site updates in $\{u\} \cup F^c$. This approximates the accelerated dynamics arbitrarily well, and contains a superset of the single-site updates of the usual Glauber dynamics. In other words, the single-site Glauber dynamics can be considered as a "censored" version of our accelerated dynamics. Theorem 4.2.4 thus completes this argument.

Let (Y_t) be the projection of the chain (X_t) onto the subgraph F. Recalling the definition of the accelerated dynamics, we see that (Y_t) is also a Markov chain, and the stationary measure ν_F for (Y_t) is the projection of μ_G to F. Furthermore, consider the subsequence (Z_t) of the chain (Y_t) obtained by skipping those times when updates occurred outside of F in (X_t). Namely, let $Z_t = Y_{K_t}$ where K_t is the t-th time that a block $\{v\} \cup F^c$ is updated in the chain (X_t). Clearly, (Z_t) is a Markov chain on the space $\{-1, 1\}^F$, where at each time a uniform vertex v from F is selected and updated according to the conditional Gibbs measure μ_G given the spins on $F \setminus \{v\}$. The stationary measure for (Z_t) is also ν_F.

Let $\mathcal{S}_t = \sum_{v \in F} Z_t(v)$ be the sum of spins over F in the chain (Z_t). It turns out that $\mathcal{S}_t$ is a distinguishing statistic and its analysis yields a lower bound on the mixing time for chain (Z_t). To this end, we need to estimate the first two moments of S_t.

Lemma 4.2.5. *Started from all-plus configuration, the sum of spins satisfies that*

$$\mathbf{E}_+(\mathcal{S}_t) \geq |F|\left(1 - \tfrac{1}{|F|}\right)^t.$$

Proof. The proof follows essentially from a coupon collecting argument. Let $(Z_t^{(+)})$ be an instance of the chain (Z_t) started at the all-plus configuration, and let (Z_t^*) be another instance of the chain (Z_t) started from ν_F. It is obvious that we can construct a monotone coupling between $(Z_t^{(+)})$ and (Z_t^*) (namely, $Z_t^{(+)} \geq Z_t^*$ for all $t \in \mathbb{N}$) such that the vertices selected for updating in both chains are always the same. Denote by $U[t]$ this (random) sequence of vertices updated up to time t. Note that Z_t^* has law ν_F, even if conditioned on the sequence $U[t]$. Recalling that $Z_t^{(+)} \geq Z_t^*$ and $\mathbf{E}_\mu \sigma(v) = 0$, we obtain that

$$\mathbf{E}_+[Z_t^{(+)}(v) \mid v \in U[t]] \geq 0.$$

It is clear that $Z_t^{(+)}(v) = 1$ if $v \notin U[t]$. Therefore,

$$\mathbf{E}_+[Z_t^{(+)}(v)] \geq \mathbf{P}(v \notin U[t]) = (1 - \tfrac{1}{|F|})^t.$$

Summing over $v \in F$ concludes the proof. ■

We next establish a contraction result for the chain (Z_t). We need the GHS inequality of [40] (see also [51] and [32]). To state this inequality, we recall the definition of the Ising model *with an external field*. Given a finite graph $G = (V, E)$ with interaction strengths $J = \{J_{uv} \geq 0 : uv \in E\}$ and external magnetic field $H = \{H_v : v \in V\}$, the probability for a configuration $\sigma \in \Omega = \{\pm 1\}^V$ is given by

$$\mu_G^H(\sigma) = \frac{1}{Z(J, H)} \exp \left(\sum_{uv \in E} J_{uv} \sigma(u) \sigma(v) + \sum_{v \in V} H(v) \sigma(v) \right), \tag{4.2.2}$$

where $Z(J, H)$ is a normalizing constant. Note that this specializes to (1.1.1) if $H \equiv 0$. When there is no ambiguity for the base graph, we sometimes drop the subscript G. We can now state the

GHS inequality [40]. For a graph $G = (V, E)$, let $\mu^H = \mu_G^H$ as above, and denote by $m_v(H) = \mathbf{E}_{\mu^H}[\sigma(v)]$ the local magnetization at vertex v. If $H_v \geq 0$ for all $v \in V$, then for any three vertices $u, v, w \in V$ (not necessarily distinct),

$$\frac{\partial^2 m_v(H)}{\partial H_u \partial H_w} \leq 0 \,.$$

The following is a consequence of the GHS inequality.

Corollary 4.2.6. *For the Ising measure μ with no external field, we have*

$$\mathbf{E}_\mu[\sigma(u) \mid v_i = 1 \text{ for all } 1 \leq i \leq k] \leq \sum_{i=1}^k \mathbf{E}_\mu[\sigma(u) \mid v_i = 1] \,.$$

Proof. The function $f(H) = m_u(H)$ satisfies $f(0) = 0$. By the GHS inequality and Claim 4.2.7 below, we obtain that for all $H, H' \in \mathbb{R}_+^n$:

$$m_u(H + H') \leq m_u(H) + m_u(H') \,. \tag{4.2.3}$$

For $1 \leq i \leq k$ and $h \geq 0$, let H_i^h be the external field taking value h on v_i and vanishing on $V \setminus \{v_i\}$. Applying the inequality (4.2.3) inductively, we deduce that

$$m_u\left(\sum_i H_i^h\right) \leq \sum_i m_u(H_i^h) \,.$$

Finally, let $h \to \infty$ and observe that $m_u(H_i^h) \to \mathbf{E}_\mu[\sigma(u) \mid \sigma(v_i) = 1]$ and $m_u(\sum_i H_i^h) \to \mathbf{E}_\mu[\sigma(u) \mid \sigma(v_i) = 1]$ for all $1 \leq i \leq k$. ■

Claim 4.2.7. *Write $\mathbb{R}_+ = [0, \infty)$ and let $f : \mathbb{R}_+^n \mapsto \mathbb{R}$ be a C^2-function such that $\frac{\partial^2 f(x)}{\partial x_i \partial x_j} \leq 0$ for all $x \in \mathbb{R}_+^n$ and $1 \leq i, j \leq n$. Then for all $x, y \in \mathbb{R}_+^n$,*

$$f(x + y) - f(x) \leq f(y) - f(0).$$

Proof. Since all the second derivatives are non-positive, $\frac{\partial f(x)}{\partial x_i}$ is decreasing in every coordinate with x for all $x \in \mathbb{R}_+^n$ and $i \leq n$. Hence, $\frac{\partial f(x)}{\partial x_i}$ is decreasing in $\mathbb{R}_+^n$. Let

$$g_x(t) = \frac{df(x + ty)}{dt} = \sum_i y_i \frac{\partial f(x)}{\partial x_i}(x + ty).$$

It follows that $g_x(t) \leq g_0(t)$ for all $x, y \in \mathbb{R}_+^n$. Integrating over $t \in [0, 1]$ yields the claim. $\blacksquare$

Lemma 4.2.8. *Suppose that $n \geq \mathrm{e}^4$. Let $(\tilde{Z}_t)$ be another instance of the chain (Z_t). Then for all starting states z_0 and $\tilde{z}_0$, there exists a coupling such that*

$$\mathbf{E}_{z_0, \tilde{z}_0} \Big[\sum_{v \in F} |Z_t(v) - \tilde{Z}_t(v)| \Big] \leq \Big(1 - \frac{1}{2|F|} \Big)^t \sum_{v \in F} |z_0(v) - \tilde{z}_0(v)|.$$

Proof. Fix $\eta, \tilde{\eta} \in \{-1, 1\}^F$ such that η and $\tilde{\eta}$ differ only at the vertex v and $\eta(v) = 1$. We consider two chains (Z_t) and $(\tilde{Z}_t)$ under monotone coupling, started from η and $\tilde{\eta}$ respectively. Let η_A be the restriction of η to A for $A \subset F$ (namely, $\eta_A \in \{-1, 1\}^A$ and $\eta_A(v) = \eta(v)$ for all $v \in A$), and write

$$\psi(u, \eta, \tilde{\eta}) = \mathbf{E}_\mu \big[\sigma(u) \mid \sigma_{F \setminus \{u\}} = \eta_{F \setminus \{u\}} \big] - \mathbf{E}_\mu \big[\sigma(u) \mid \sigma_{F \setminus \{u\}} = \tilde{\eta}_{F \setminus \{u\}} \big].$$

By the monotone property and symmetry of the Ising model,

$$\psi(u, \eta, \tilde{\eta}) \leq \mathbf{E}_\mu[\sigma(u) \mid \sigma_{F \setminus \{u\}} = +] - \mathbf{E}_\mu[\sigma(u) \mid \sigma_{F \setminus \{u\}} = -]$$
$$= 2\mathbf{E}_\mu[\sigma(u) \mid \sigma_{F \setminus \{u\}} = +].$$

By symmetry, we see that $\mathbf{E}(\sigma(u) \mid \sigma(w) = 1) = -\mathbf{E}(\sigma(u) \mid \sigma(w) = -1)$ and $\mathbf{E}(\sigma(u)) = 0$. Thus, $\mathrm{Cov}(\sigma(u), \sigma(w)) = \mathbf{E}(\sigma(u) \mid \sigma(w) = 1)$. Combined with Corollary 4.2.6, it yields that

$$\psi(u, \eta, \tilde{\eta}) \leq 2 \sum_{w \in F \setminus \{u\}} \mathbf{E}_\mu[\sigma(u) \mid \sigma(w) = 1] = 2 \sum_{w \in F \setminus \{u\}} \mathrm{Cov}(\sigma(u), \sigma(w)),$$

Recalling the non-negative correlations between the spins, we deduce that under the mono-

tone coupling

$$\mathbf{E}_{\eta,\tilde{\eta}}\Big[\frac{1}{2}\sum_{v'\in F}|Z_1(v') - \tilde{Z}_1(v')|\Big] = 1 - \frac{1}{|F|} + \frac{1}{2|F|}\sum_{u\in F\backslash\{v\}}\psi(u,\eta,\tilde{\eta})$$

$$\leq 1 - \frac{1}{|F|} + \frac{1}{|F|}\sum_{u\in F\backslash\{v\}}\sum_{w\in F\backslash\{u\}}\mathrm{Cov}(\sigma(u),\sigma(w)).$$

By Lemma 4.2.2, we get that for $n \geq \mathrm{e}^4$,

$$\mathbf{E}_{\eta,\tilde{\eta}}\Big[\frac{1}{2}\sum_{v'\in F}|Z_1(v') - \tilde{Z}_1(v')|\Big] \leq 1 - \frac{1}{|F|} + \frac{2}{|F|\log n} \leq 1 - \frac{1}{2|F|}.$$

Using the triangle inequality and recursion, we conclude the proof. ∎

From the contraction result, we can derive the uniform variance bound on $\mathcal{S}_t$. This type of argument appeared in [53] (see Lemma 2.4) when (Z_t) is a one dimensional chain. The argument naturally extends to multi-dimensional case and we include the proof for completeness.

Lemma 4.2.9. *Let (Z_t) and $(\tilde{Z}_t)$ be two instances of a Markov chain taking values in $\mathbb{R}^n$. Assume that for some $\rho < 1$ and all initial states z_0 and $\tilde{z}_0$, there exists a coupling satisfying*

$$\mathbf{E}_{z_0,\tilde{z}_0}\Big[\sum_i|Z_t(i) - \tilde{Z}_t(i)|\Big] \leq \rho^t\sum_i|z_0(i) - \tilde{z}_0(i)|,$$

where we used the convention that $z(i)$ stands for the i-th coordinate of z for $z \in \mathbb{R}^n$. Furthermore, suppose that $\sum_i|Z_t(i) - Z_{t-1}(i)| \leq R$ for all t. Then for any $t \in \mathbb{N}$ and starting state $z \in \mathbb{R}^n$,

$$\mathrm{Var}_z\Big(\sum_i Z_t(i)\Big) \leq \frac{2}{1-\rho^2}R^2.$$

Proof. Let Z_t and Z'_t be two independent instances of the chain both started from z. Defining $Q_t = \sum_i Z_t(i)$ and $Q'_t = \sum_i Z'_t(i)$, we obtain that

$$\Big|\mathbf{E}_z[Q_t \mid Z_1 = z_1] - \mathbf{E}_z[Q'_t|Z'_1 = z'_1]\Big| = \Big|\mathbf{E}_{z_1}[Q_{t-1}] - \mathbf{E}_{z'_1}[Q'_{t-1}]\Big|$$

$$\leq \rho^{t-1}\sum_i|z_1(i) - z'_1(i)| \leq 2\rho^{t-1}R,$$

for all possible choices of z_1 and z'_1. It follows that for any starting state z

$$\mathrm{Var}_z(\mathbf{E}_z[Q_t \mid Z_1]) = \tfrac{1}{2}\mathbf{E}_z\Big[\big(\mathbf{E}_{Z_1}[Q_{t-1}] - \mathbf{E}_{Z'_1}[Q'_{t-1}]\big)^2\Big] \leq 2(\rho^{t-1}R)^2.$$

Therefore, by the total variance formula, we obtain that for all z

$$\mathrm{Var}_z(Q_t) = \mathrm{Var}_z(\mathbf{E}_z[Q_t \mid Z_1]) + \mathbf{E}_z[\mathrm{Var}_z(Q_t \mid Z_1)] \leq 2(\rho^{t-1}R)^2 + \nu_{t-1}\,,$$

where $\nu_t := \max_z \mathrm{Var}_z(Q_t)$. Thus $\nu_t \leq 2(\rho^{t-1}R)^2 + \nu_{t-1}$, whence

$$\nu_t \leq \sum_{i=1}^{t}(\nu_i - \nu_{i-1}) \leq \sum_{i=1}^{t} 2\rho^{2(t-1)}R^2 \leq \frac{2R^2}{1-\rho^2}\,,$$

completing the proof. $\blacksquare$

Combining the above two lemmas gives the following variance bound (note that in our case $R = 2$ and $\rho = 1 - \frac{1}{2|F|}$, so $1 - \rho^2 \geq \frac{1}{2|F|}$).

Lemma 4.2.10. *For all t and starting position z, we have $\mathrm{Var}_z(\mathcal{S}_t) \leq 16|F|$.*

We can now derive a lower bound on the mixing time for the chain (Z_t).

Lemma 4.2.11. *The chain (Z_t) has a mixing time $t_{\mathrm{MIX}}^+ \geq \frac{1}{2}|F|\log|F| - 20|F|$.*

Proof. Let $(Z_t^{(+)})$ be an instance of the dynamics (Z_t) started from the all-plus configuration and let $Z^* \in \{-1,1\}^F$ be distributed as ν_F. Write

$$T_0 = \tfrac{1}{2}|F|\log|F| - 20|F|\,.$$

It suffices to prove that

$$d_{\mathrm{TV}}(\mathcal{S}_{T_0}^{(+)}, \mathcal{S}^*) \geq \tfrac{1}{4}\,, \tag{4.2.4}$$

where $\mathcal{S}_{T_0}^{(+)} = \sum_{v \in F} Z_{T_0}^{(+)}(v)$ as before and $\mathcal{S}^* = \sum_{v \in F} Z^*(v)$ be the sum of spins in stationary distribution. To this end, notice that by Lemmas 4.2.5 and 4.2.10:

$$\mathbf{E}_+(\mathcal{S}_{T_0}^{(+)}) \geq \mathrm{e}^{20+o(1)}\sqrt{|F|} \quad \text{and} \quad \mathrm{Var}_+(\mathcal{S}_{T_0}^{(+)}) \leq 16|F|\,.$$

An application of Chebyshev's inequality gives that for large enough n

$$\mathbf{P}_+(\mathcal{S}_{T_0}^{(+)} \leq \mathrm{e}^{10}\sqrt{|F|}) \leq \frac{16|F|}{(\mathrm{e}^{20+o(1)} - \mathrm{e}^{10})\sqrt{|F|})^2} \leq \frac{1}{4}\,. \tag{4.2.5}$$

On the other hand, it is clear by symmetry that $\mathbf{E}_{\nu_F}\mathcal{S}^* = 0$. Moreover, since Lemma 4.2.10 holds for all t, taking $t \to \infty$ gives that $\mathrm{Var}_{\nu_F}\mathcal{S}^* \leq 16|F|$. Applying Chebyshev's inequality

again, we deduce that

$$\mathbf{P}_{\nu_F}(\mathcal{S}^* \geq \mathrm{e}^{10}\sqrt{|F|}) \leq \frac{16|F|}{(\mathrm{e}^{10}\sqrt{|F|})^2} \leq \frac{1}{4}\,.$$

Combining the above inequality with (4.2.5) and the fact that

$$d_{\mathrm{TV}}(\mathcal{S}_{T_0}^{(+)}, \mathcal{S}^*) \geq 1 - \mathbf{P}_+(\mathcal{S}_{T_0}^{(+)} \leq \mathrm{e}^{10}\sqrt{|F|}) - \mathbf{P}_\mu(\mathcal{S}^* \geq \mathrm{e}^{10}\sqrt{|F|})\,,$$

we conclude that (4.2.5) indeed holds (with room to spare), as required. ∎

We are now ready to derive Theorem 9. Observe that the dynamics (Y_t) is a lazy version of the dynamics (Z_t). Consider an instance (Y_t^+) of the dynamics (Y_t) started from the all-plus configuration and let $Y^* \in \{-1, 1\}^F$ be distributed according to the stationary distribution ν_F. Let $\mathcal{S}_t^{(+)}$ and $\mathcal{S}^*$ again be the sum of spins over F, but with respect to the chain $(Y_t^{(+)})$ and the variable Y^* respectively. Write

$$T = \frac{n}{|F|}\left(\frac{1}{2}|F|\log|F| - 40|F|\right),$$

and let N_T be the number of steps in $[1, T]$ where a block of the form $\{v\} \cup F$ is selected to update in the chain $(Y_t^{(+)})$. By Chebyshev's inequality,

$$\mathbf{P}(N_T \geq \tfrac{1}{2}|F|\log|F| - 20|F|) \leq \frac{T|F|/n}{(20|F|)^2} = o(1)\,.$$

Repeating the arguments in the proof of Lemma 4.2.11, we deduce that for all $t \leq T_0 = \frac{1}{2}|F|\log|F| - 20|F|$, we have

$$\mathbf{P}_+(\mathcal{S}_t^{(+)} \leq \mathrm{e}^{10}\sqrt{|F|}) \leq \tfrac{1}{4}\,.$$

Therefore

$$\|\mathbf{P}_+(Y_T^{(+)} \in \cdot) - \nu_F\|_{\mathrm{TV}} \geq 1 - \mathbf{P}(N_T \geq T_0) - \mathbf{P}_{\mu_Y}(\mathcal{S}^* \geq \mathrm{e}^{10}\sqrt{|F|})$$
$$- \mathbf{P}_+\big(\mathcal{S}_T^{(+)} \leq \mathrm{e}^{10}\sqrt{|F|} \mid N_T \leq T_0\big)\,.$$

Altogether, we have that

$$\|\mathbf{P}_+(Y_T^{(+)} \in \cdot) - \nu_F\|_{\mathrm{TV}} \geq \tfrac{1}{2} + o(1) \geq \tfrac{1}{4}\,,$$

and hence that
$$t_{\text{MIX}}^{+,Y} \geq T \geq \tfrac{1+o(1)}{4} n \log n \,,$$

where $t_{\text{MIX}}^{+,Y}$ refers to the mixing time for chain $(Y_t^{(+)})$. Since the chain (Y_t) is a projection of the chain (X_t), it follows that the mixing time for the chain (X_t) satisfies $t_{\text{MIX}}^{+,X} \geq (1/4 + o(1))n \log n$. Combining this bound with Theorem 4.2.4 (see the discussion following the statement of the theorem), we conclude that the Glauber dynamics started with the all-plus configuration has mixing time $t_{\text{MIX}}^{+} \geq (1/4 + o(1))n \log n$. $\blacksquare$

Remark. The analysis naturally extends to the continuous-time Glauber dynamics, where each site is associated with an independent Poisson clock of unit rate determining the update times of this site as above (note that the continuous dynamics is $|V|$ times faster than the discrete dynamics). We can use similar arguments to these used above to handel the laziness in the transition from the chain (Z_t) to the chain (Y_t). Namely, we could condition on the number of updates up to time t and then repeat the above arguments to establish that $t_{\text{MIX}}^{+} \geq (1/4 + o(1)) \log n$ in the continuous-time case.

Remark. We believe that Theorem 9 should have analogues (with t_{MIX} in place of t_{MIX}^{+}) for the Ising model with arbitrary magnetic field, as well as for the Potts model and proper colorings. The first of these may be accessible to the methods of this chapter, but the other two models need new ideas.

Remark. For the Ising model in a box of $\mathbb{Z}^d$ at high temperature, it is well known that the mixing time is $\Theta(n \log n)$. Recently, the sharp asymptotics (the so-called cutoff phenomenon) was established by Lubetzky and Sly [59].

Part II

Random walk on random graphs

Chapter 5

Mixing time for the random walk on near-supercritical random graphs

5.1 Preliminaries

5.1.1 Cores and kernels

The *k-core* of a graph G, denoted by $G^{(k)}$, is the maximum subgraph $H \subset G$ where every vertex has degree at least k. It is well known (and easy to see) that this subgraph is unique, and can be obtained by repeatedly deleting any vertex whose degree is smaller than k (at an arbitrary order).

We call a path $\mathcal{P} = v_0, v_1, \ldots, v_k$ for $k > 1$ (i.e., a sequence of vertices with $v_i v_{i+1}$ an edge for each i) a *2-path* if and only if v_i has degree 2 for all $i = 1, \ldots, k-1$ (while the endpoints v_0, v_k may have degree larger than 2, and possibly $v_0 = v_k$).

The *kernel* $\mathcal{K}$ of G is obtained by taking its 2-core $G^{(2)}$ minus its disjoint cycles, then repeatedly contracting all 2-paths (replacing each by a single edge). Note that, by definition, the degree of every vertex in $\mathcal{K}$ is at least 3.

5.1.2 Structure of the supercritical giant component

The key to our analysis of the random walk on the giant component $\mathcal{C}_1$ is the following result from [22]. This theorem completely characterizes the structure of $\mathcal{C}_1$, by reducing it to a tractable contiguous model $\tilde{\mathcal{C}}_1$.

Theorem 5.1.1. [22] *Let $\mathcal{C}_1$ be the largest component of $\mathcal{G}(n, p)$ for $p = \frac{1+\varepsilon}{n}$, where $\varepsilon^3 n \to \infty$ and $\varepsilon \to 0$. Let $\mu < 1$ denote the conjugate of $1 + \varepsilon$, that is, $\mu e^{-\mu} = (1 + \varepsilon)e^{-(1+\varepsilon)}$. Then $\mathcal{C}_1$ is contiguous to the following model $\tilde{\mathcal{C}}_1$:*

1. *Let $\Lambda \sim \mathcal{N}\left(1 + \varepsilon - \mu, \frac{1}{\varepsilon n}\right)$ and assign i.i.d. variables $D_u \sim \text{Poisson}(\Lambda)$ $(u \in [n])$ to the vertices, conditioned that $\sum D_u \mathbf{1}_{\{D_u \geq 3\}}$ is even. Let*

$$N_k = \#\{u : D_u = k\} \quad and \quad N = \sum_{k \geq 3} N_k.$$

Select a random multigraph $\mathcal{K}$ on N vertices, uniformly among all multigraphs with N_k vertices of degree k for $k \geq 3$.

2. *Replace the edges of $\mathcal{K}$ by paths of lengths i.i.d. $\text{Geom}(1 - \mu)$.*

3. *Attach an independent $\text{Poisson}(\mu)$-Galton-Watson tree to each vertex.*

That is, $\mathbf{P}(\tilde{\mathcal{C}}_1 \in \mathcal{A}) \to 0$ implies $\mathbf{P}(\mathcal{C}_1 \in \mathcal{A}) \to 0$ for any set of graphs $\mathcal{A}$ that is closed under graph-isomorphism.

In the above, a $\text{Poisson}(\mu)$-Galton-Watson tree is the family tree of a Galton-Watson branching process with offspring distribution $\text{Poisson}(\mu)$. We will use the abbreviation $\text{PGW}(\mu)$-tree for this object. A multigraph is the generalization of a simple graph permitting multiple edges and loops.

Note that conditioning on $\sum D_u \mathbf{1}_{\{D_u \geq 3\}}$ being even does not pose a problem, as one can easily use rejection sampling. The 3 steps in the description of $\tilde{\mathcal{C}}_1$ correspond to constructing its kernel $\mathcal{K}$ (Step 1), expanding $\mathcal{K}$ into the 2-core $\tilde{\mathcal{C}}_1^{(2)}$ (Step 2), and finally attaching trees to it to obtain $\tilde{\mathcal{C}}_1$ (Step 3).

Further observe that $N_k \asymp \varepsilon^k n$ for any fixed $k \geq 2$, and so in the special case where $\varepsilon = o(n^{-1/4})$ w.h.p. we have $D_u \in \{0, 1, 2, 3\}$ for all $u \in [n]$, and the kernel $\mathcal{K}$ is simply a uniform 3-regular multigraph.

Combining the above description of the giant component with standard tools in the study of random graphs with given degree-sequences, one can easily read off useful geometric properties of the kernel. This is demonstrated by the following lemma of [22], for which we require a few definitions: For a vertex v in G let $d_G(v)$ denote its degree and for a subset of vertices S let

$$d_G(S) := \sum_{v \in S} d_G(v)$$

denote the sum of the degrees of its vertices (also referred to as the *volume* of S in G). The *isoperimetric number* of a graph G is defined to be

$$i(G) := \min\left\{\frac{e(S, S^c)}{d_G(S)} \; : \; S \subset V(G), \, d_G(S) \leq e(G)\right\},$$

where $e(S,T)$ denotes the number of edges between S and T while $e(G)$ is the total number of edges in G.

Lemma 5.1.2 ([22, Lemma 3.5]). *Let $\mathcal{K}$ be the kernel of the largest component $\mathcal{C}_1$ of $\mathcal{G}(n,p)$ for $p = \frac{1+\varepsilon}{n}$, where $\varepsilon^3 n \to \infty$ and $\varepsilon \to 0$. Then w.h.p.,*

$$|\mathcal{K}| = \left(\tfrac{4}{3} + o(1)\right)\varepsilon^3 n\,, \quad e(\mathcal{K}) = (2 + o(1))\varepsilon^3 n\,,$$

and $i(\mathcal{K}) \geq \alpha$ for some absolute constant $\alpha > 0$.

5.1.3 Notions of mixing of the random walk

For any two distributions φ, ψ on V, the *total-variation distance* of φ and ψ is defined as

$$\|\varphi - \psi\|_{\mathrm{TV}} := \sup_{S \subset V} |\varphi(S) - \psi(S)| = \frac{1}{2}\sum_{v \in V}|\varphi(v) - \psi(v)|\,.$$

Let (S_t) denote the lazy random walk on G, i.e., the Markov chain which at each step holds its position with probability $\frac{1}{2}$ and otherwise moves to a uniformly chosen neighbor. This is an aperiodic and irreducible Markov chain, whose stationary distribution π is given by

$$\pi(x) = d_G(x)/2|E|\,.$$

We next define two notions of measuring the distance of an ergodic Markov chain (S_t), defined on a state-set V, from its stationary distribution π.

Let $0 < \delta < 1$. The (worst-case) *total-variation mixing time* of (S_t) with parameter δ, denoted by $t_{\mathrm{MIX}}(\delta)$, is defined to be

$$t_{\mathrm{MIX}}(\delta) := \min\left\{t : \max_{v \in V}\|\mathbf{P}_v(S_t \in \cdot) - \pi\|_{\mathrm{TV}} \leq \delta\right\},$$

where $\mathbf{P}_v$ denotes the probability given that $S_0 = v$.

The *Cesàro mixing time* (also known as the approximate uniform mixing time) of (S_t) with parameter δ, denoted by $\tilde{t}_{\mathrm{MIX}}(\delta)$, is defined as

$$\tilde{t}_{\mathrm{MIX}}(\delta) = \min\left\{t : \max_{v \in V}\left\|\pi - \frac{1}{t}\sum_{i=0}^{t-1}\mathbf{P}_v(S_i \in \cdot)\right\|_{\mathrm{TV}} \leq \delta\right\}.$$

When discussing the order of the mixing-time it is customary to choose $\delta = \frac{1}{4}$, in which case we will use the abbreviations $t_{\mathrm{MIX}} = t_{\mathrm{MIX}}(\frac{1}{4})$ and $\tilde{t}_{\mathrm{MIX}} = \tilde{t}_{\mathrm{MIX}}(\frac{1}{4})$.

By results of [4] and [57] (see also [58]), the mixing time and the Cesàro mixing time have the same order for lazy reversible Markov chains (i.e., discrete-time chains whose holding probability in each state is at least $\frac{1}{2}$), as formulated by the following theorem.

Theorem 5.1.3. *Every lazy reversible Markov chain satisfies*

$$c_1\, \tilde{t}_{\mathrm{MIX}}(\tfrac{1}{4}) \le t_{\mathrm{MIX}}(\tfrac{1}{4}) \le c_2\, \tilde{t}_{\mathrm{MIX}}(\tfrac{1}{4})$$

for some absolute constants $c_1, c_2 > 0$.

Proof. The first inequality is straightforward and does not require laziness or reversibility. We include its proof for completeness. Notice that

$$\left\| \pi - \frac{1}{t} \sum_{i=0}^{t-1} \mathbf{P}_v(S_i \in \cdot) \right\|_{\mathrm{TV}} \le \frac{1}{8} + \frac{1}{t} \sum_{i=t/8}^{t-1} \left\| \pi - \mathbf{P}_v(S_i \in \cdot) \right\|_{\mathrm{TV}}$$

$$\le \frac{1}{8} + \left\| \pi - \mathbf{P}_v(S_{t/8} \in \cdot) \right\|_{\mathrm{TV}},$$

where we used the fact that $\|\pi - \mathbf{P}_v(S_t \in \cdot)\|$ is decreasing in t. Taking $t = 8t_{\mathrm{MIX}}(\frac{1}{8})$, we obtain that $\tilde{t}_{\mathrm{MIX}}(\frac{1}{4}) \le 8t_{\mathrm{MIX}}(\frac{1}{8})$ and conclude the proof of the first inequality using the well-known fact that $t_{\mathrm{MIX}}(\frac{1}{8}) \le 4t_{\mathrm{MIX}}(\frac{1}{4})$.

The second inequality of the theorem is significantly more involved: By combining [57, Theorem 5.4] (for a stronger version, see [58, Theorem 4.22]) and [4, Theorem C], it follows that the order of the Cesàro mixing time can be bounded by that of the mixing time for the corresponding continuous-time Markov chain. Now, using a well-known fact that the mixing time for the lazy Markov chain and the continuous-time chain have the same order (see, e.g., [54, Theorem 20.3]), the proof is concluded. ∎

Let Γ be a stopping rule (a randomized stopping time) for (S_t). That is, $\Gamma : G \times \Omega \to \mathbb{N}$ for some probability space Ω, such that $\Gamma(\cdot, \omega)$ is a stopping time for every $\omega \in \Omega$. Let $\sigma^\Gamma := \mathbf{P}_\sigma(S_\Gamma \in \cdot)$ when σ is a distribution on V.

Let σ, ν be two distributions on V. Note that there is always a stopping rule Γ such that $\sigma^\Gamma = \nu$, e.g., draw a vertex z according to ν and stop when reaching z. The *access time* from σ to ν, denoted by $H(\sigma, \nu)$, is the minimum expected number of steps over all such stopping rules:

$$H(\sigma, \nu) := \min_{\Gamma\, :\, \sigma^\Gamma = \nu} \mathbf{E}\Gamma.$$

It is easy to verify that $H(\sigma, \nu) = 0$ iff $\sigma = \nu$ and that $H(\cdot, \cdot)$ satisfies the triangle-inequality, however it is not necessarily symmetric.

The *approximate forget time* of G with parameter $0 < \delta < 1$ is defined by

$$\mathcal{F}_\delta = \min_\varphi \max_\sigma \min_{\nu \,:\, \|\nu - \varphi\|_{\mathrm{TV}} \leq \delta} H(\sigma, \nu) \,. \tag{5.1.1}$$

Combining Theorem 3.2 and Corollary 5.4 in [58], one immediately obtains that the approximate forget time and the Cesàro mixing time have the same order, as stated in the following theorem.

Theorem 5.1.4. *Every reversible Markov chain satisfies*

$$c_1 \, \mathcal{F}_{1/4} \leq \tilde{t}_{\mathrm{MIX}}(\tfrac{1}{4}) \leq c_2 \, \mathcal{F}_{1/4}$$

for some absolute constants $c_1, c_2 > 0$.

5.1.4   Conductance and mixing

Let $P = (p_{x,y})_{x,y}$ be the transition kernel of an irreducible, reversible and aperiodic Markov chain on Ω with stationary distribution π. For $S \subset \Omega$, define the *conductance* of the set S to be

$$\Phi(S) := \frac{\sum_{x \in S, y \notin S} \pi(x) p_{x,y}}{\pi(S) \pi(\Omega \setminus S)} \,.$$

We define Φ, the conductance of the chain, by $\Phi := \min\{\Phi(S) \,:\, \pi(S) \leq \tfrac{1}{2}\}$ (In the special case of a lazy random walk on a connected regular graph, this quantity is similar to the isoperimetric number of the graph, defined earlier). A well-known result of Jerrum and Sinclair [47] states that t_{MIX} is of order at most $\Phi^{-2} \log \pi_{\min}^{-1}$, where $\pi_{\min} = \min_{x \in \Omega} \pi(x)$. This bound was fine-tuned by Lovász and Kannan [56] to exploit settings where the conductance of the average set S plays a dominant role (rather than the worst set). For our upper bound of the mixing time on the random walk on the 2-core, we will use an enhanced version of the latter bound (namely, Theorem 5.2.6) due to Fountoulakis and Reed [37].

5.1.5   Edge set notations

Throughout the chapter we will use the following notations, which will be handy when moving between the kernel and 2-core.

For $S \subset G$, let $E_G(S)$ denote the set of edges in the induced subgraph of G on S, and let $\partial_G S$ denote the edges between S and its complement $S^c := V(G) \setminus S$. Let

$$\bar{E}_G(S) := E_G(S) \cup \partial_G(S) \,,$$

and define $e_G(S) := |E_G(S)|$. We omit the subscript G whenever its identity is made clear from the context.

If $\mathcal{K}$ is the kernel in the model $\tilde{\mathcal{C}}_1$ and $\mathcal{H}$ is its 2-core, let

$$E_{\mathcal{H}}^{\star} : 2^{E(\mathcal{K})} \to 2^{E(\mathcal{H})}$$

be the operator which takes a subset of edges $T \subset E(\mathcal{K})$ and outputs the edges lying on their corresponding 2-paths in $\mathcal{H}$. For $S \subset V(\mathcal{K})$, we let

$$E_{\mathcal{H}}^{\star}(S) := E_{\mathcal{H}}^{\star}(E_{\mathcal{K}}(S)) , \quad \bar{E}_{\mathcal{H}}^{\star}(S) := E_{\mathcal{H}}^{\star}(\bar{E}_{\mathcal{K}}(S)) .$$

5.2 Random walk on the 2-core

In this section we analyze the properties of the random walk on the 2-core $\tilde{\mathcal{C}}_1^{(2)}$.

5.2.1 Mixing time of the 2-core

By the definition of our new model $\tilde{\mathcal{C}}_1$, we can study the 2-core $\mathcal{C}_1^{(2)}$ via the well-known configuration model (see, e.g., [10] for further details on this method). To simplify the notation, we let $\mathcal{H}$ denote the 2-core of $\tilde{\mathcal{C}}_1$ throughout this section.

The main goal of the subsection is to establish the mixing time of the lazy random walk on $\mathcal{H}$, as stated by the following theorem.

Theorem 5.2.1. *With high probability, the lazy random walk on $\mathcal{H}$ has a Cesàro mixing time $\tilde{t}_{\mathrm{MIX}}$ of order $\varepsilon^{-2} \log^2(\varepsilon^3 n)$. Consequently, w.h.p. it also satisfies $t_{\mathrm{MIX}} \asymp \varepsilon^{-2} \log^2(\varepsilon^3 n)$.*

We will use a result of Fountoulakis and Reed [38], which bounds the mixing time in terms of the isoperimetric profile of the graph (measuring the expansion of sets of various volumes). As a first step in obtaining this data for the supercritical 2-core $\mathcal{H}$, the next lemma will show that a small subset of the kernel, $S \subset \mathcal{K}$, cannot have too many edges in $\bar{E}_{\mathcal{H}}(S)$.

Lemma 5.2.2. *For $v \in \mathcal{K}$, define*

$$\mathfrak{C}_{v,K} := \{ S \ni v : \ |S| = K \ and \ S \ is \ a \ connected \ subgraph \ of \ \mathcal{K} \ \} .$$

The following holds w.h.p. for every $v \in \mathcal{K}$, integer K and $S \in \mathfrak{C}_{v,K}$:

1. $|\mathfrak{C}_{v,K}| \leq \exp \left[5(K \vee \log(\varepsilon^3 n)) \right]$.

2. $d_{\mathcal{K}}(S) \leq 30(K \vee \log(\varepsilon^3 n))$.

Proof. By definition, $\Lambda = (2 + o(1))\varepsilon$ w.h.p., thus standard concentration arguments imply that the following holds w.h.p.:

$$N_3 = \left(\tfrac{4}{3} + o(1)\right)\varepsilon^3 n \quad \text{and} \quad N_k \leq \frac{(3\varepsilon)^k \log(1/\varepsilon)}{k!} n \text{ for } k \geq 4 \,. \tag{5.2.1}$$

Assume that the above indeed holds, and notice that the lemma trivially holds when $K \geq \varepsilon^3 n$. We may therefore further assume that $K \leq \varepsilon^3 n$.

Consider the following exploration process, starting from the vertex v. Initialize S to be $\{v\}$, and mark $v_1 = v$. At time $i \geq 1$, we explore the neighborhood of v_i (unless $|S| < i$), and for each its neighbors that does not already belong to S, we toss a fair coin to decide whether or not to insert it to S. Newly inserted vertices are labeled according to the order of their arrival; that is, if $|S| = k$ prior to the insertion, we give the new vertex the label v_{k+1}. Finally, if $|S| < i$ at time i then we stop the exploration process.

Let X_i denote the degree of the vertex v_i in the above defined process. In order to stochastically dominate X_i from above, observe that the worst case occurs when each of the vertices in $v_1, \ldots, v_{i-1}$ has degree 3. With this observation in mind, let A be a set consisting of $N_3 - K$ vertices of degree 3 and N_k vertices k (for $k \geq 4$). Sample a vertex proportional to the degree from A and let Y denote its degree. Clearly, $X_i \preceq Y_i$, where Y_i are independent variables distributed as Y, and so

$$d_{\mathcal{K}}(S) \preceq \sum_{i=1}^{K} Y_i \,. \tag{5.2.2}$$

By the definition of our exploration process,

$$|\mathfrak{C}_{v,K}| \preceq \sum_{\ell_1 + \cdots + \ell_K = K} \prod_{i=1}^{K} \binom{Y_i}{\ell_i} \,.$$

We can now deduce that

$$\mathbf{E}|\mathfrak{C}_{v,K}| \leq \mathbf{E}\left[\sum_{\ell_1 + \cdots + \ell_K = K} \prod_{i=1}^{K} \binom{Y_i}{\ell_i} \right] = \sum_{\ell_1 + \cdots + \ell_K = K} \prod_{i=1}^{K} \mathbf{E}\left[\binom{Y_i}{\ell_i} \right] \,. \tag{5.2.3}$$

For all $i \geq 4$, we have

$$\mathbf{P}(Y = i) \leq 27i \frac{(3\varepsilon)^{i-3} \log(1/\varepsilon)}{i!} = 27 \frac{(3\varepsilon)^{i-3} \log(1/\varepsilon)}{(i-1)!}$$

and therefore, for sufficiently large n (recall that $\varepsilon = o(1)$),

$$\mathbf{E}\left[\binom{Y}{k}\right] \leq \binom{3}{k} + \sum_{i \geq 4} \binom{i}{k} \cdot 27 \frac{(3\varepsilon)^{i-3} \log(1/\varepsilon)}{(i-1)!} \leq \frac{7}{k!} \text{ for all } k\,,$$

Altogether,

$$\mathbf{E}|\mathfrak{C}_{v,K}| \leq 7^K \sum_{\ell_1 + \cdots + \ell_K = K} \prod_{i=1}^{K} \frac{1}{\ell_i!}\,. \tag{5.2.4}$$

The next simple claim will provide a bound on the sum in the last expression.

Claim 5.2.3. *The function* $f(n) = \sum_{\ell_1 + \cdots + \ell_n = n} \prod_{k=1}^{n} \frac{1}{\ell_k!}$ *satisfies* $f(n) \leq \mathrm{e}^n$.

Proof. The proof is by induction. For $n = 1$, the claim trivially holds. Assuming the hypothesis is valid for $n \leq m$, we get

$$f(m+1) = \sum_{k=0}^{m+1} \frac{1}{k!} f(m-k) \leq \sum_{k=0}^{m+1} \frac{\mathrm{e}^{m-k}}{k!} \leq \mathrm{e}^m \sum_{k=0}^{m+1} \frac{1}{k!} \leq \mathrm{e}^{m+1}\,,$$

as required. ∎

Plugging the above estimate into (5.2.4), we conclude that $\mathbf{E}|\mathfrak{C}_{v,K}| \leq (7\mathrm{e})^K$. Now, Markov's inequality, together with a union bound over all the vertices in the kernel $\mathcal{K}$ yield the Part (1) of the lemma.

For Part (2), notice that for any sufficiently large n,

$$\mathbf{E}\mathrm{e}^Y \leq \mathrm{e}^3 + \sum_{i \geq 4} \mathrm{e}^i 27 i \frac{(3\varepsilon)^{i-3} \log(1/\varepsilon)}{i!} \leq 25\,,$$

Therefore, (5.2.2) gives that

$$\mathbf{P}\left(d_{\mathcal{K}}(S) \geq 30\left(K \vee \log(\varepsilon^3 n)\right)\right) \leq \exp\left[-5\left(K \vee \log(\varepsilon^3 n)\right)\right]\,.$$

At this point, the proof is concluded by a union bound over $\mathfrak{C}_{v,K}$ for all $v \in \mathcal{K}$ and $K \leq \varepsilon^3 n$, using the upper bound we have already derived for $|\mathfrak{C}_{v,K}|$ in the Part (1) of the lemma. ∎

Lemma 5.2.4. *Let* $\mathcal{L} \subset E(\mathcal{K})$ *be the set of loops in the kernel. With high probability, every subset of vertices* $S \subset \mathcal{K}$ *forming a connected subgraph of* $\mathcal{K}$ *satisfies* $|\bar{E}_{\mathcal{H}}^{\star}(S)| \leq (100/\varepsilon)\left(|S| \vee \log(\varepsilon^3 n)\right)$, *and every subset* T *of* $\frac{1}{20}\varepsilon^3 n$ *edges in* $\mathcal{K}$ *satisfies* $|E_{\mathcal{H}}^{\star}(T) \cup E_{\mathcal{H}}^{\star}(\mathcal{L})| \leq \frac{3}{4}\varepsilon^2 n$.

Proof. Assume that the events given in Parts (1),(2) of Lemma 5.2.2 hold. Further note that, by definition of the model $\tilde{\mathcal{C}}_1$, a standard application of CLT yields that w.h.p.

$$|\mathcal{K}| = (\tfrac{4}{3} + o(1))\varepsilon^3 n\,, \quad e(\mathcal{H}) = (2 + o(1))\varepsilon^2 n\,, \quad e(\mathcal{K}) = (2 + o(1))\varepsilon^3 n\,.$$

By Part (2) of that lemma, $d_{\mathcal{K}}(S) \leq 30\,(|S| \vee \log(\varepsilon^3 n))$ holds simultaneously for every connected set S, hence there are at most this many edges in $\bar{E}_{\mathcal{K}}(S)$.

Let $S \subset \mathcal{K}$ be a connected set of size $|S| = s$, and let

$$K = K(s) = s \vee \log(\varepsilon^3 n)\,.$$

Recalling our definition of the graph $\mathcal{H}$, we deduce that

$$\left|\bar{E}_{\mathcal{H}}^{\star}(S)\right| \preceq \sum_{i=1}^{30K} Z_i\,,$$

where Z_i are i.i.d. Geometric random variables with mean $\frac{1}{1-\mu}$. It is well known that the moment-generating function of such variables is given by

$$\mathbf{E}(e^{tZ_1}) = \frac{(1-\mu)e^t}{1 - \mu e^t}\,.$$

Setting $t = \varepsilon/2$ and recalling that $\mu = 1 - (1+o(1))\varepsilon$, we get that $\mathbf{E}(e^{\frac{\varepsilon}{2}Z_1}) \leq e$ for sufficiently large n (recall that $\varepsilon = o(1)$). Therefore, we obtain that for the above mentioned S,

$$\mathbf{P}\left(|\bar{E}_{\mathcal{H}}^{\star}(S)| \geq (100/\varepsilon)K\right) \leq \frac{\exp(30K)}{\exp(\frac{\varepsilon}{2}(100/\varepsilon)K)} = e^{-20K}\,.$$

By Part (1) of Lemma 5.2.2, there are at most $(\tfrac{4}{3} + o(1))\varepsilon^3 n \exp(5K)$ connected sets of size s. Taking a union bound over the $(\tfrac{4}{3} + o(1))\varepsilon^3 n$ values of s establishes that the statement of the lemma holds except with probability

$$\left(\tfrac{4}{3} + o(1)\right) \varepsilon^3 n \sum_s e^{-20K(s)} e^{5K(s)} \leq \left(\tfrac{16}{9} + o(1)\right) (\varepsilon^3 n)^{-13} = o(1)\,,$$

completing the proof of the statement on all connected subsets $S \subset \mathcal{K}$.

Next, if T contains t edges in $\mathcal{K}$, then the number of corresponding edges in $\mathcal{H}$ is again stochastically dominated by a sum of i.i.d. geometric variables $\{Z_i\}$ as above. Hence, by the same argument, the probability that there exists a set $T \subset E(\mathcal{K})$ of $\alpha \varepsilon^3 n$ edges in $\mathcal{K}$, which

expands to at least $\beta\varepsilon^2 n$ edges in $\mathcal{H}$ for some $0 < \alpha < \frac{1}{2}$ and $0 < \beta < 1$, is at most

$$\binom{(2 + o(1))\varepsilon^3 n}{\alpha\varepsilon^3 n} \frac{\mathrm{e}^{\alpha\varepsilon^3 n}}{\mathrm{e}^{(\varepsilon/2)\beta\varepsilon^2 n}} \leq \exp\left[\left(2H\left(\tfrac{\alpha}{2}\right) + \alpha - \tfrac{\beta}{2} + o(1)\right)\varepsilon^3 n\right]$$

(using the well-known fact that $\sum_{i \leq \lambda m} \binom{m}{i} \leq \exp[H(\lambda)m]$ where $H(x)$ is the entropy function $H(x) := -x\log x - (1 - x)\log(1 - x)$). It is now easy to verify that a choice of $\alpha = \frac{1}{20}$ and $\beta = \frac{2}{3}$ in the last expression yields a term that tends to 0 as $n \to \infty$.

It remains to bound $|\mathcal{L}|$. This will follow from a bound on the number of loops in $\mathcal{K}$. Let $u \in \mathcal{K}$ be a kernel vertex, and recall that its degree D_u is distributed as an independent $\big(\mathrm{Poisson}(\Lambda)\,\big|\,\cdot \geq 3\big)$, where $\Lambda = (2 + o(1))\varepsilon$ with high probability. The expected number of loops that u obtains in a random realization of the degree sequence (via the configuration model) is clearly at most D_u^2/D, where $D = (4 + o(1))\varepsilon^3 n$ is the total of the kernel degrees. Therefore,

$$\mathbf{E}|\mathcal{L}| \leq \left(\tfrac{4}{3} + o(1)\right)\varepsilon^3 n \cdot (1/D)\mathbf{E}[D_u^2] = O(1)\,,$$

and so $\mathbf{E}|E_{\mathcal{H}}^\star(\mathcal{L})| = O(1/\varepsilon)$. The contribution of $|E_{\mathcal{H}}^\star(\mathcal{L})|$ is thus easily absorbed w.h.p. when increasing β from $\frac{2}{3}$ to $\frac{3}{4}$, completing the proof. $\blacksquare$

Lemma 5.2.5. *There exists an absolute constant $\iota > 0$ so that w.h.p. every connected set $S \subset \mathcal{H}$ with $(200/\varepsilon)\log(\varepsilon^3 n) \leq d_{\mathcal{H}}(S) \leq e(\mathcal{H})$ satisfies that $|\partial_{\mathcal{H}}S|\,/\,d_{\mathcal{H}}(S) \geq \iota\varepsilon$.*

Proof. Let $S \subset \mathcal{H}$ be as above, and write $S_{\mathcal{K}} = S \cap \mathcal{K}$. Observe that $S_{\mathcal{K}}$ is connected (if nonempty). Furthermore, since $d_{\mathcal{H}}(S) \geq (200/\varepsilon)\log(\varepsilon^3 n)$ whereas the longest 2-path in $\mathcal{H}$ contains $(1 + o(1))(1/\varepsilon)\log(\varepsilon^3 n)$ edges w.h.p., we may assume that $S_{\mathcal{K}}$ is indeed nonempty.

Next, clearly $|\partial_{\mathcal{H}}S| \geq |\partial_{\mathcal{K}}S_{\mathcal{K}}|$ (as each edge in the boundary of $S_{\mathcal{K}}$ translates into a 2-path in $\mathcal{H}$ with precisely one endpoint in S), while $|\bar{E}_{\mathcal{H}}(S)| \leq |\bar{E}_{\mathcal{H}}^\star(S_{\mathcal{K}})|$ (any $e \in \bar{E}_{\mathcal{H}}(S)$ belongs to some 2-path $\mathcal{P}_e$, which is necessarily incident to some $v \in S_{\mathcal{K}}$ as, crucially, $S_{\mathcal{K}}$ is nonempty. Hence, the edge corresponding to $\mathcal{P}_e$ belongs to $\bar{E}_{\mathcal{K}}(S_{\mathcal{K}})$, and so $e \in \bar{E}_{\mathcal{H}}^\star(S_{\mathcal{K}})$). Therefore, using the fact that $d_{\mathcal{H}}(S) \leq 2|\bar{E}_{\mathcal{H}}(S)|$,

$$\frac{|\partial_{\mathcal{H}}S|}{d_{\mathcal{H}}(S)} \geq \frac{|\partial_{\mathcal{K}}S_{\mathcal{K}}|}{2|\bar{E}_{\mathcal{H}}^\star(S_{\mathcal{K}})|} = \frac{|\partial_{\mathcal{K}}S_{\mathcal{K}}|}{2|\bar{E}_{\mathcal{K}}(S_{\mathcal{K}})|} \cdot \frac{|\bar{E}_{\mathcal{K}}(S_{\mathcal{K}})|}{|\bar{E}_{\mathcal{H}}^\star(S_{\mathcal{K}})|}\,. \tag{5.2.5}$$

Assume that the events stated in Lemma 5.2.4 hold. Since the assumption on $d_{\mathcal{H}}(S_{\mathcal{K}})$ gives that $|\bar{E}_{\mathcal{H}}^\star(S_{\mathcal{K}})| \geq (100/\varepsilon)\log(\varepsilon^3 n)$, we deduce that necessarily

$$|S_{\mathcal{K}}| \geq (\varepsilon/100)|\bar{E}_{\mathcal{H}}^\star(S_{\mathcal{K}})|\,,$$

and thus (since $S_\mathcal{K}$ is connected)

$$|\bar{E}_\mathcal{K}(S_\mathcal{K})| \geq |E_\mathcal{K}(S_\mathcal{K})| \geq (\varepsilon/100)|\bar{E}_\mathcal{H}^\star(S_\mathcal{K})| - 1 \,. \tag{5.2.6}$$

Now,

$$d_\mathcal{H}(S) \leq e(\mathcal{H}) = (2 + o(1))\varepsilon^2 n \,,$$

and since $d_\mathcal{H}(S) = 2|E_\mathcal{H}(S)| + |\partial_\mathcal{H} S|$ we have $|E_\mathcal{H}(S)| \leq (1 + o(1))\varepsilon^2 n$. In particular, $|E(\mathcal{H}) \setminus E_\mathcal{H}(S)| \geq \frac{3}{4}\varepsilon^2 n$ for sufficiently large n.

At the same time, if $\mathcal{L}$ is the set of all loops in $\mathcal{K}$ and $T = \bar{E}_\mathcal{K}(\mathcal{K} \setminus S_\mathcal{K})$, then clearly $E_\mathcal{H}^\star(T) \cup E_\mathcal{H}^\star(\mathcal{L})$ is a superset of $E(\mathcal{H}) \setminus E_\mathcal{H}(S)$. Therefore, Lemma 5.2.4 yields that $|T| \geq \frac{1}{20}\varepsilon^3 n$. Since $d_\mathcal{K}(S_\mathcal{K}) \leq 2e(\mathcal{K}) = (4 + o(1))\varepsilon^3 n$, we get

$$d_\mathcal{K}(\mathcal{K} \setminus S_\mathcal{K}) \geq |T| \geq \frac{\varepsilon^3 n}{20} \geq \frac{1 + o(1)}{80} d_\mathcal{K}(S_\mathcal{K}) \,.$$

At this point, by Lemma 5.1.2 there exists $\alpha > 0$ such that w.h.p. for any such above mentioned subset S:

$$|\partial_\mathcal{K} S_\mathcal{K}| \geq \alpha \left(d_\mathcal{K}(S_\mathcal{K}) \wedge d_\mathcal{K}(\mathcal{K} \setminus S_\mathcal{K}) \right) \geq \frac{\alpha + o(1)}{80} d_\mathcal{K}(S_\mathcal{K}) \,. \tag{5.2.7}$$

Plugging (5.2.6),(5.2.7) into (5.2.5), we conclude that the lemma holds for any sufficiently large n with, say, $\iota = \frac{1}{2} \cdot 10^{-4}\alpha$. ∎

We are now ready to establish the upper bound on the mixing time for the random walk on $\mathcal{H}$.

Proof of Theorem 5.2.1. We will apply the following recent result of [37], which bounds the mixing time of a lazy chain in terms of its isoperimetric profile (a fine-tuned version of the Lovász-Kannan [56] bound on the mixing time in terms of the average conductance).

Theorem 5.2.6 ([37]). *Let $P = (p_{x,y})$ be the transition kernel of an irreducible, reversible and aperiodic Markov chain on Ω with stationary distribution π. Let $\pi_{\min} = \min_{x \in \Omega} \pi(x)$ and for $p > \pi_{\min}$, let*

$$\Phi(p) := \min\{\Phi(S) : S \text{ is connected and } p/2 \leq \pi(S) \leq p\} \,,$$

and $\Phi(p) = 1$ *if there is no such S. Then for some absolute constant $C > 0$,*

$$\tilde{t}_{\text{MIX}} \leq C \sum_{j=1}^{\lceil \log \pi_{\min}^{-1} \rceil} \Phi^{-2}(2^{-j}) \,.$$

In our case, the P is the transition kernel of the lazy random walk on $\mathcal{H}$. By definition, if $S \subset \mathcal{H}$ and $d_{\mathcal{H}}(x)$ denotes the degree of $x \in \mathcal{H}$, then

$$\pi_{\mathcal{H}}(x) = \frac{d_{\mathcal{H}}(x)}{2e(\mathcal{H})} \,, \quad p_{x,y} = \frac{1}{2d_{\mathcal{H}}(x)} \,, \quad \pi_{\mathcal{H}}(S) = \frac{d_{\mathcal{H}}(S)}{2e(\mathcal{H})} \,,$$

and so $\Phi(S) \geq \frac{1}{2}|\partial_{\mathcal{H}}S|/d_{\mathcal{H}}(S)$. Recall that w.h.p. $e(\mathcal{H}) = (2 + o(1))\varepsilon^2 n$. Under this assumption, for any $p \geq 120\frac{\log(\varepsilon^3 n)}{\varepsilon^3 n}$ and connected subset $S \subset \mathcal{H}$ satisfying $\pi_{\mathcal{H}}(S) \geq p/2$,

$$d_{\mathcal{H}}(S) = 2\pi_{\mathcal{H}}(S)e(\mathcal{H}) \geq (200/\varepsilon)\log(\varepsilon^3 n) \,.$$

Therefore, by Lemma 5.2.5, w.h.p.

$$\Phi(p) \geq \tfrac{1}{2}\iota\varepsilon \quad \text{for all} \quad 120\frac{\log(\varepsilon^3 n)}{\varepsilon^3 n} \leq p \leq \tfrac{1}{2} \,. \tag{5.2.8}$$

Set

$$j^* = \max\left\{ j : 2^{-j} \geq 120\frac{\log(\varepsilon^3 n)}{\varepsilon^3 n} \right\} \,.$$

It is clear that $j^* = O(\log(\varepsilon^3 n))$ and (5.2.8) can be translated into

$$\Phi(2^{-j}) \geq \tfrac{1}{2}\iota\varepsilon, \text{ for all } 1 \leq j \leq j^* \,. \tag{5.2.9}$$

On the other hand, if $\pi_{\mathcal{H}}(S) \leq p < 1$ then $d_{\mathcal{H}}(S) \leq 2pe(\mathcal{H})$ while $|\partial_{\mathcal{H}}S| \geq 1$ (as $\mathcal{H}$ is connected), and so the inequality $\Phi(S) \geq \frac{1}{2}|\partial_{\mathcal{H}}S|/d_{\mathcal{H}}(S)$ gives $\Phi(S) \geq 1/(4pe(\mathcal{H}))$. Substituting $p = 2^{-j}$ with $j \leq \lceil \log \pi_{\min}^{-1} \rceil$ we have

$$\Phi(2^{-j}) \geq \frac{2^{j-2}}{e(\mathcal{H})} \geq \frac{2^j}{10\varepsilon^2 n} \tag{5.2.10}$$

(where the last inequality holds for large n). Combining (5.2.9) and (5.2.10) together, we

now apply Theorem 5.2.6 to conclude that there exists a constant $C > 0$ such that, w.h.p.,

$$\tilde{t}_{\text{MIX}} \leq C \sum_{j=1}^{\lceil \log \pi_{\min}^{-1} \rceil} \frac{1}{\Phi^2(2^{-j})} = C \left[\sum_{j=1}^{j^*} \frac{1}{\Phi^2(2^{-j})} + \sum_{j=j^*}^{\lceil \log \pi_{\min}^{-1} \rceil} \frac{1}{\Phi^2(2^{-j})} \right]$$
$$\leq C \left(j^* (\tfrac{1}{2} \iota \varepsilon)^{-2} + 2(10\varepsilon^2 n \cdot 2^{-j^*})^2 \right) = O(\varepsilon^{-2} \log^2(\varepsilon^3 n)) \, ,$$

where the last inequality follows by our choice of j^*.

The lower bound on the mixing time follows immediately from the fact that, by the definition of $\tilde{\mathcal{C}}_1$, w.h.p. there exists a 2-path in $\mathcal{H}$ whose length is $(1 - o(1))(1/\varepsilon) \log(\varepsilon^3 n)$ (see [22, Corollary 1]). ∎

5.2.2 Local times for the random walk on the 2-core

In order to extend the mixing time from the 2-core $\mathcal{H}$ to the giant component, we need to prove the following proposition.

Proposition 5.2.7. *Let $N_{v,s}$ be the local time induced by the lazy random walk (W_t) on $\mathcal{H}$ to the vertex v up to time s, i.e., $\#\{0 \leq t \leq s : W_t = v\}$. Then there exists some $C > 0$ such that, w.h.p., for all $s > 0$ and any $u, v \in \mathcal{H}$,*

$$\mathbf{E}_u[N_{v,s}] \leq C \frac{\varepsilon s}{\log(\varepsilon^3 n)} + (150/\varepsilon) \log(\varepsilon^3 n) \, .$$

In order to prove Proposition 5.2.7, we wish to show that with positive probability the random walk W_t will take an excursion in a long 2-path before returning to v. Consider some $v \in \mathcal{K}$ (we will later extend this analysis to the vertices in $\mathcal{H} \setminus \mathcal{K}$, i.e., those vertices lying on 2-paths). We point out that proving this statement is simpler in case $D_v = O(1)$, and most of the technical challenge lies in the possibility that D_v is unbounded. In order to treat this point, we first show that the neighbors of vertex v in the kernel are, in some sense, distant apart.

Lemma 5.2.8. *For $v \in \mathcal{K}$ let $\mathcal{N}_v$ denote the set of neighbors of v in the kernel $\mathcal{K}$. Then w.h.p., for every $v \in \mathcal{K}$ there exists a collection of disjoint connected subsets $\{B_w(v) \subset \mathcal{K} : w \in \mathcal{N}_v\}$, such that for all $w \in \mathcal{N}_v$,*

$$|B_w| = \lceil (\varepsilon^3 n)^{1/5} \rceil \quad and \quad \text{diam}(B_w) \leq \tfrac{1}{2} \log(\varepsilon^3 n) \, .$$

Proof. We may again assume (5.2.1) and furthermore, that

$$3 \leq D_v \leq \log(\varepsilon^3 n) \text{ for all } v \in \mathcal{K} \, .$$

Let $v \in \mathcal{K}$. We construct the connected sets B_w while we reveal the structure of the kernel $\mathcal{K}$ via the configuration model, as follows: Process the vertices $w \in \mathcal{N}_v$ sequentially according to some arbitrary order. When processing such a vertex w, we expose the ball (according to the graph metric) about it, excluding v and any vertices that were already accounted for, until its size reaches $\lceil (\varepsilon^3 n)^{1/5} \rceil$ (or until no additional new vertices can be added).

It is clear from the definition that the B_w's are indeed disjoint and connected, and it remains to prove that each B_w satisfies $|B_w| = \lceil (\varepsilon^3 n)^{1/5} \rceil$ and $\mathrm{diam}(B_w) \leq \log(\varepsilon^3 n)$.

Let R denote the tree-excess of the (connected) subset $\{v\} \cup \bigcup_w B_w$ once the process is concluded. We claim that w.h.p. $R \leq 1$. To see this, first observe that at any point in the above process, the sum of degrees of all the vertices that were already exposed (including v and $\mathcal{N}_v$) is at most

$$\lceil (\varepsilon^3 n)^{1/5} \rceil \log^2(\varepsilon^3 n) = (\varepsilon^3 n)^{1/5 + o(1)}.$$

Hence, by the definition of the configuration model (which draws a new half-edge between w and some other vertex proportional to its degree), $R \preceq Z$ where Z is a binomial variable $\mathrm{Bin}\left((\varepsilon^3 n)^{1/5 + o(1)}, (\varepsilon^3 n)^{-4/5 + o(1)}\right)$. This gives

$$\mathbf{P}(R \geq 2) = (\varepsilon^3 n)^{-6/5 + o(1)}.$$

In particular, since $D_w \geq 3$ for any $w \in \mathcal{K}$, this implies that we never fail to grow B_w to size $(\varepsilon^3 n)^{1/5}$, and that the diameter of each B_w is at most that of a binary tree (possibly plus $R \leq 1$), i.e., for any large n,

$$\mathrm{diam}(B_w) \leq \tfrac{1}{5} \log_2(\varepsilon^3 n) + 2 \leq \tfrac{1}{2} \log(\varepsilon^3 n).$$

A simple union bound over $v \in \mathcal{K}$ now completes the proof. ∎

We distinguish the following subset of the edges of the kernel, whose paths are suitably long:

$$\mathcal{E} := \left\{ e \in E(\mathcal{K}) : |\mathcal{P}_e| \geq \frac{1}{20\varepsilon} \log(\varepsilon^3 n) \right\},$$

where $\mathcal{P}_e$ is the 2-path in $\mathcal{H}$ that corresponds to the edge $e \in E(\mathcal{K})$. Further define $\mathcal{Q} \subset 2^{\mathcal{K}}$ to be all the subsets of vertices of $\mathcal{K}$ whose induced subgraph contains an edge from $\mathcal{E}$:

$$\mathcal{Q} := \{ S \subset \mathcal{K} : E_{\mathcal{K}}(S) \cap \mathcal{E} \neq \emptyset \}.$$

For each $e \in \mathcal{K}$, we define the median of its 2-path, denoted by $\mathrm{med}(\mathcal{P}_e)$, in the obvious manner: It is the vertex $w \in \mathcal{P}_e$ whose distance from the two endpoints is the same, up to at most 1 (whenever there are two choices for this w, pick one arbitrarily). Now, for each

$v \in \mathcal{H}$ let

$$\mathcal{E}_v := \{\operatorname{med}(\mathcal{P}_e) \ : \ e \in \mathcal{E} \ , \ v \notin \mathcal{P}_e\} \,.$$

The next lemma provides a lower bound on the effective conductance between a vertex v in the 2-core and its corresponding above defined set $\mathcal{E}_v$. See, e.g., [63] for further details on conductances/resistances.

Lemma 5.2.9. *Let $C_{\mathrm{eff}}(v \leftrightarrow \mathcal{E}_v)$ be the effective conductance between a vertex $v \in \mathcal{H}$ and the set $\mathcal{E}_v$. With high probability, for any $v \in \mathcal{H}$,*

$$C_{\mathrm{eff}}(v \leftrightarrow \mathcal{E}_v)/D_v \geq \varepsilon/(100 \log(\varepsilon^3 n)) \,.$$

Proof. In order to bound the effective conductance, we need to prove that for any $v \in \mathcal{K}$, there exist D_v disjoint paths of length at most $(100/\varepsilon) \log(\varepsilon^3 n)$ leading to the set $\mathcal{E}_v$. By Lemmas 5.2.4 and 5.2.8, it suffices to prove that w.h.p. for any $v \in \mathcal{K}$ and $w \in \mathcal{N}_v$, we have that $E(B_w) \cap \mathcal{E} \neq \emptyset$, where $\mathcal{N}_v$ and B_w are defined as in Lemma 5.2.8 (in this case, the path from v to some $e \in \mathcal{E}$ within B_w will have length at most $\frac{1}{2} \log(\varepsilon^3 n)$ in $\mathcal{K}$, and its length will not be exceed $(100/\varepsilon) \log(\varepsilon^3 n)$ after being expanded in the 2-core).

Notice that if Y is the geometric variable $\mathrm{Geom}(1 - \mu)$ then

$$\mathbf{P}\big(Y \geq \tfrac{1}{10\varepsilon} \log(\varepsilon^3 n)\big) = \mu^{\frac{1}{10\varepsilon} \log(\varepsilon^3 n)} \geq (\varepsilon^3 n)^{-1/10 + o(1)} \,.$$

Therefore, by the independence of the lengths of the 2-paths and the fact that $|B_w| = \big\lceil (\varepsilon^3 n)^{1/5} \big\rceil$, we obtain that

$$\mathbf{P}\left(E(B_w) \cap \mathcal{E} = \emptyset\right) \leq \big(1 - (\varepsilon^3 n)^{-1/10 + o(1)}\big)^{(\varepsilon^3 n)^{1/5}} \leq \mathrm{e}^{-(\varepsilon^3 n)^{1/10 - o(1)}} \,.$$

At this point, a union bound shows that the probability that for some $v \in \mathcal{K}$ there exists some $w \in \mathcal{N}_v$, such that $E(B_w)$ does not intersect $\mathcal{E}$, is at most

$$\big(\tfrac{4}{3} + o(1)\big) \varepsilon^3 n \cdot \log(\varepsilon^3 n) \cdot \mathrm{e}^{-(\varepsilon^3 n)^{1/10 - o(1)}} = o(1) \,. \qquad \blacksquare$$

We are ready to prove the main result of this subsection, Proposition 5.2.7, which bounds the local times induced by the random walk on the 2-core.

Proof of Proposition 5.2.7. For some vertex $v \in \mathcal{H}$ and subset $A \subset \mathcal{H}$, let

$$\tau_v^+ := \min\{t > 0 : W_t = v\} \,, \quad \tau_A := \min\{t : W_t \in A\} \,.$$

It is well-known (see, e.g., [63, equation (2.4)]) that the effective conductance has the fol-

lowing form:

$$\mathbf{P}_v(\tau_A < \tau_v^+) = \frac{C_{\mathrm{eff}}(v \leftrightarrow A)}{D_v} \, .$$

Combined with Lemma 5.2.9, it follows that

$$\mathbf{P}_v(\tau_{\mathcal{E}_v} < \tau_v^+) = \frac{C_{\mathrm{eff}}(v \leftrightarrow \mathcal{E}_v)}{D_v} \geq \varepsilon/\left(100 \log(\varepsilon^3 n)\right) \, .$$

On the other hand, for any $v \in \mathcal{H}$, by definition $w \in \mathcal{E}_v$ is the median of some 2-path, which does not contain v and has length at least $\frac{1}{20\varepsilon} \log(\varepsilon^3 n)$. Hence, by well-known properties of hitting times for the simple random walk on the integers, there exists some absolute constant $c > 0$ such that for any $v \in \mathcal{H}$ and $w \in \mathcal{E}_v$:

$$\mathbf{P}_w\left(\tau_v^+ \geq c\varepsilon^{-2} \log^2(\varepsilon^3 n)\right) \geq \mathbf{P}_w\left(\tau_{\mathcal{K}} \geq c\varepsilon^{-2} \log^2(\varepsilon^3 n)\right) \geq \tfrac{2}{3} \, ,$$

Altogether, we conclude that

$$\mathbf{P}_v\left(\tau_v^+ \geq c\varepsilon^{-2} \log^2(\varepsilon^3 n)\right) \geq \mathbf{P}_v\left(\tau_{\mathcal{E}_v} < \tau_v^+\right) \min_{w \in \mathcal{E}_v} \left\{\mathbf{P}_w(\tau_v^+ \geq c\varepsilon^{-2} \log^2(\varepsilon^3 n))\right\}$$

$$\geq \varepsilon/\left(150 \log(\varepsilon^3 n)\right) \, .$$

Setting $t_c = c\varepsilon^{-2} \log^2(\varepsilon^3 n)$, we can rewrite the above as

$$\mathbf{P}_v(N_{v,t_c} \geq 2) \leq 1 - \varepsilon/\left(150 \log(\varepsilon^3 n)\right) \, .$$

By the strong Markovian property (i.e., $(W_{\tau_v^+ + t})$ is a Markov chain with the same transition kernel of (W_t)), we deduce that

$$\mathbf{P}(N_{v,t_c} \geq k) \leq \left[1 - \varepsilon/\left(150 \log(\varepsilon^3 n)\right)\right]^{k-1} \, ,$$

and hence

$$\mathbf{E} N_{v,t_c} \leq (150/\varepsilon) \log(\varepsilon^3 n) \, .$$

The proof is completed by observing that $\mathbf{E}_v(N_{v,s}) \leq \lceil s/t_c \rceil \mathbf{E}_v N_{v,t_c}$ and that $\mathbf{E}_u N_{v,s} \leq \mathbf{E}_v N_{v,s}$ for any u. ∎

5.3 Mixing on the giant component

In this section, we prove Theorem 1, which establishes the order of the mixing time of the lazy random walk on the supercritical $\mathcal{C}_1$.

5.3.1 Controlling the attached Poisson Galton-Watson trees

So far, we have established that w.h.p. the mixing time of the lazy random walk on the 2-core $\tilde{\mathcal{C}}_1^{(2)}$ has order $\varepsilon^{-2} \log^2(\varepsilon^3 n)$. To derive the mixing time for $\tilde{\mathcal{C}}_1$ based on that estimate, we need to consider the delays due to the excursions the random walk makes in the attached trees. As we will later see, these delays will be upper bounded by a certain a linear combination of the sizes of the trees (with weights determined by the random walk on the 2-core). The following lemma will play a role in estimating this expression.

Lemma 5.3.1. *Let $\{\mathcal{T}_i\}$ be independent $\mathrm{PGW}(\mu)$-trees. For any two constants $C_1, C_2 > 0$ there exists some constant $C > 0$ such that the following holds: If $\{a_i\}_{i=1}^m$ is a sequence of positive reals satisfying*

$$\sum_{i=1}^m a_i \leq C_1 \varepsilon^{-2} \log^2(\varepsilon^3 n), \tag{5.3.1}$$

$$\max_{1 \leq i \leq m} a_i \leq C_2 \varepsilon^{-1} \log(\varepsilon^3 n), \tag{5.3.2}$$

then

$$\mathbf{P}\Big(\sum_{i=1}^m a_i |\mathcal{T}_i| \geq C \varepsilon^{-3} \log^2(\varepsilon^3 n) \Big) \leq (\varepsilon^3 n)^{-2}.$$

Proof. It is well-known (see, e.g., [76]) that the size of a Poisson(γ)-Galton-Watson tree $\mathcal{T}$ follows a Borel(γ) distribution, namely,

$$\mathbf{P}(|\mathcal{T}| = k) = \frac{k^{k-1}}{\gamma k!} (\gamma \mathrm{e}^{-\gamma})^k. \tag{5.3.3}$$

The following is a well-known (and easy) estimate on the size of a PGW-tree; we include its proof for completeness.

Claim 5.3.2. *Let $0 < \gamma < 1$, and let $\mathcal{T}$ be a $\mathrm{PGW}(\gamma)$-tree. Then*

$$\mathbf{E}|\mathcal{T}| = \frac{1}{1-\gamma}, \qquad \mathrm{Var}(|\mathcal{T}|) = \frac{\gamma}{(1-\gamma)^3}.$$

Proof. For $k = 0, 1, \ldots$, let L_k be the number of vertices in the k-th level of the tree $\mathcal{T}$. Clearly, $\mathbf{E}L_k = \gamma^k$, and so $\mathbf{E}|\mathcal{T}| = \mathbf{E}\sum_k L_k = \frac{1}{1-\gamma}$.

By the total-variance formula,

$$\mathrm{Var}(L_i) = \mathrm{Var}\big(\mathbf{E}\big(L_i \,\big|\, L_{i-1}\big)\big) + \mathbf{E}\big(\mathrm{Var}\big(L_i \,\big|\, L_{i-1}\big)\big)$$
$$= \gamma^2 \mathrm{Var}(L_{i-1}) + \gamma \mathbf{E}L_{i-1} = \gamma^2 \mathrm{Var}(L_{i-1}) + \gamma^i.$$

By induction,

$$\mathrm{Var}(L_i) = \sum_{k=i}^{2i-1} \gamma^k = \gamma^i \frac{1-\gamma^i}{1-\gamma}. \tag{5.3.4}$$

We next turn to the covariance of L_i, L_j for $i \leq j$:

$$\mathrm{Cov}(L_i, L_j) = \mathbf{E}[L_i L_j] - \mathbf{E} L_i \mathbf{E} L_j = \gamma^{j-i} \mathbf{E} L_i^2 - \gamma^{i+j}$$

$$= \gamma^{j-i} \mathrm{Var}(L_i) = \gamma^j \frac{1-\gamma^i}{1-\gamma}.$$

Summing over the variances and covariances of the L_i's, we deduce that

$$\mathrm{Var}(|\mathcal{T}|) = 2 \sum_{i=0}^{\infty} \sum_{j=i}^{\infty} \gamma^j \frac{1-\gamma^i}{1-\gamma} - \sum_{i=0}^{\infty} \gamma^i \frac{1-\gamma^i}{1-\gamma} = \frac{\gamma}{(1-\gamma)^3}. \qquad \blacksquare$$

We need the next lemma to bound the tail probability for $\sum a_i |\mathcal{T}_i|$.

Lemma 5.3.3 ([49, Corollary 4.2]). *Let $X_1, \ldots, X_m$ be independent r.v.'s with $\mathbf{E}[X_i] = \mu_i$. Suppose there are b_i, d_i and ξ_0 such that $\mathrm{Var}(X_i) \leq b_i$, and*

$$\left| \mathbf{E}\left[(X_i - \mu_i)^3 e^{\xi(X_i - \mu_i)} \right] \right| \leq d_i \quad \text{for all } 0 \leq |\xi| \leq \xi_0.$$

If $\delta \xi_0 \sum_{i=1}^m d_i \leq \sum_{i=1}^m b_i$ for some $0 < \delta \leq 1$, then for all $\Delta > 0$,

$$\mathbf{P}\left(\left| \sum_{i=1}^m X_i - \sum_{i=1}^m \mu_i \right| \geq \Delta \right) \leq \exp\left(-\frac{1}{3} \min\left\{ \delta \xi_0 \Delta, \frac{\Delta^2}{\sum_{i=1}^m b_i} \right\} \right).$$

Let $T_i = |\mathcal{T}_i|$ and $X_i = a_i T_i$ for $i \in [m]$. Claim 5.3.2 gives that

$$\mu_i = \mathbf{E} X_i = a_i/(1-\mu).$$

Now set

$$\xi_0 = \varepsilon^3/(10 C_2 \log(\varepsilon^3 n)).$$

For any $|\xi| \leq \xi_0$, we have $a_i |\xi| \leq \varepsilon^2/10$ by the assumption (5.3.2), and so

$$\left| \mathbf{E}\left[(X_i - \mu_i)^3 e^{\xi(X_i - \mu_i)} \right] \right| = a_i^3 \left| \mathbf{E}\left[\left(T_i - \frac{1}{1-\mu} \right)^3 e^{\xi a_i (T_i - \frac{1}{1-\mu})} \right] \right|$$

$$\leq a_i^3 \mathbf{E}\left[(1-\mu)^{-3} \mathbf{1}_{\{T_i < (1-\mu)^{-1}\}} \right] + a_i^3 \mathbf{E}\left[T_i^3 e^{\xi a_i T_i} \mathbf{1}_{\{T_i \geq (1-\mu)^{-1}\}} \right]$$

$$\leq a_i^3 (1-\mu)^{-3} + a_i^3 \mathbf{E}\left[T_i^3 \exp(\varepsilon^2 T_i/10) \right]. \tag{5.3.5}$$

Recalling the law of T_i given by (5.3.3), we obtain that

$$\mathbf{E}\left(T_i^3 \exp(\varepsilon^2 T_i/10)\right) = \sum_{k=1}^{\infty} \frac{k^{k-1}}{\mu k!}(\mu e^{-\mu})^k k^3 e^{\varepsilon^2 k/10} .$$

Using Stirling's formula, we obtain that for some absolute constant $c > 1$,

$$\mathbf{E}\left(T_i^3 \exp(\varepsilon^2 T_i/10)\right) \le c \sum_{k=1}^{\infty} \frac{k^{k-1}(\mu e^{-\mu})^k}{\mu(k/e)^k \sqrt{k}} k^3 e^{\varepsilon^2 k/10}$$

$$= \frac{c}{\mu} \sum_{k=1}^{\infty} k^{3/2}(\mu e^{1-\mu})^k e^{\varepsilon^2 k/10} . \tag{5.3.6}$$

Recalling that $\mu = 1 - \varepsilon + \frac{2}{3}\varepsilon^2 + O(\varepsilon^3)$ and using the fact that $1 - x \le e^{-x-x^2/2}$ for $x \ge 0$, we get that for sufficiently large n (and hence small enough ε),

$$\mu e^{1-\mu} = (1 - (1 - \mu)) e^{1-\mu} \le \exp\left(-\tfrac{1}{2}\varepsilon^2 + O(\varepsilon^3)\right) \le e^{-\varepsilon^2/3} . \tag{5.3.7}$$

Plugging the above estimate into (5.3.6), we obtain that for large n,

$$\mathbf{E}\left[T_i^3 \exp(\varepsilon^2 T_i/10)\right] \le 2c \sum_{k=1}^{\infty} k^{3/2} e^{-\varepsilon^2 k/6} \le 4c \int_0^{\infty} x^{3/2} e^{-\varepsilon^2 x/6} \, dx$$

$$\le 400c\varepsilon^{-5} \int_0^{\infty} x^{3/2} e^{-x} \, dx = 300\sqrt{\pi} c\varepsilon^{-5} .$$

Going back to (5.3.5), we get that for some absolute $c' > 1$ and any large n,

$$\left|\mathbf{E}\left[(X_i - \mu_i)^3 e^{\xi(X_i - \mu_i)}\right]\right| \le a_i^3 \left(2\varepsilon^{-3} + c'\varepsilon^{-5}\right) \le a_i \cdot 2c'C_2^2\varepsilon^{-7} \log^2(\varepsilon^3 n) := d_i ,$$

where the second inequality used (5.3.2).

By Claim 5.3.2, it follows that for large enough n,

$$\operatorname{Var}(X_i) = a_i^2 \operatorname{Var}(T_i) = a_i^2 \frac{\mu}{(1-\mu)^3} \le 2a_i^2\varepsilon^{-3} \le a_i \cdot 2C_2\varepsilon^{-4} \log(\varepsilon^3 n) := b_i .$$

Since $\sum_i d_i = (c'C_2\varepsilon^{-3} \log(\varepsilon^3 n)) \sum_i b_i$, by setting $\delta = 1$ (and recalling our choice of ξ_0) we get

$$\delta\xi_0 \sum_{i=1}^{m} d_i = \frac{\delta c'}{10} \sum_i b_i \le \sum_{i=1}^{m} b_i .$$

We have thus established the conditions for Lemma 5.3.3, and it remains to select Δ. For a choice of $\Delta = (60C_2 \vee \sqrt{12C_1C_2})\varepsilon^{-3}\log^2(\varepsilon^3 n)$, by definition of ξ_0 and the b_i's we have

$$\xi_0\Delta \geq 6\log(\varepsilon^3 n)\,,$$
$$\Delta^2/\textstyle\sum_i b_i \geq 6C_1\varepsilon^{-2}\log^3(\varepsilon^3 n)/\textstyle\sum_i a_i \geq 6\log(\varepsilon^3 n)\,,$$

where the last inequality relied on (5.3.1). Hence, an application of Lemma 5.3.3 gives that for large enough n,

$$\mathbf{P}\Big(\textstyle\sum_i a_i T_i - \textstyle\sum_i \mu_i \geq \Delta\Big) \leq (\varepsilon^3 n)^{-2}\,.$$

Finally, by (5.3.1) and using the fact that $1 - \mu \geq \varepsilon/2$ for any large n, we have $\sum_i \mu_i = (1-\mu)^{-1}\sum_i a_i \leq 2C_1\varepsilon^{-3}\log^2(\varepsilon^3 n)$. The proof of Lemma 5.3.1 is thus concluded by choosing $C = 2C_1 + (60C_2 \vee \sqrt{12C_1C_2})$. ∎

To bound the time it takes the random walk to exit from an attached PGW-tree (and enter the 2-core), we will need to control the diameter and volume of such a tree. The following simple lemma of [23] gives an estimate on the diameter of a PGW-tree:

Lemma 5.3.4 ([23, Lemma 3.2]). *Let $\mathcal{T}$ be a $\mathrm{PGW}(\mu)$-tree and L_k be its k-th level of vertices. Then $\mathbf{P}(L_k \neq \emptyset) \asymp \varepsilon e^{-k(\varepsilon+O(\varepsilon^2))}$ for any $k \geq 1/\varepsilon$.*

The next lemma gives a bound on the volume of a PGW-tree:

Lemma 5.3.5. *Let $\mathcal{T}$ be a $\mathrm{PGW}(\mu)$-tree. Then*

$$\mathbf{P}(|\mathcal{T}| \geq 6\varepsilon^{-2}\log(\varepsilon^3 n)) = o(\varepsilon(\varepsilon^3 n)^{-2})\,.$$

Proof. Recalling (5.3.3) and applying Stirling's formula, we obtain that for any $s > 0$,

$$\mathbf{P}(|\mathcal{T}| \geq s) = \sum_{k \geq s} \frac{k^{k-1}}{\mu k!}(\mu e^{-\mu})^k \asymp \sum_{k \geq s} \frac{(\mu e^{1-\mu})^k}{k^{3/2}}\,. \tag{5.3.8}$$

Write $r = \log(\varepsilon^3 n)$. By estimate (5.3.7), we now get that for large enough n,

$$\sum_{k \geq 6\varepsilon^{-2}r} k^{-3/2}(\mu e^{1-\mu})^k \leq \sum_{k \geq 6\varepsilon^{-2}r} k^{-3/2}e^{-\varepsilon^2 k/3} = O(e^{-2r}\varepsilon/\sqrt{r})\,,$$

and combined with (5.3.8) this concludes the proof. ∎

Finally, for the lower bound, we will need to show that w.h.p. one of the attached PGW-trees in $\tilde{\mathcal{C}}_1$ is suitably large, as we next describe. For a rooted tree $\mathcal{T}$, let L_k be its k-th level

of vertices and $\mathcal{T}_v$ be its entire subtree rooted at v. Define the event

$$A_{r,s}(\mathcal{T}) := \{\exists v \in L_r \text{ such that } |\mathcal{T}_v| \geq s \}.$$

The next lemma gives a bound on the probability of this event when $\mathcal{T}$ is a $\mathrm{PGW}(\mu)$-tree.

Lemma 5.3.6. *Let $\mathcal{T}$ be a $\mathrm{PGW}(\mu)$-tree and take $r = \lceil \frac{1}{8}\varepsilon^{-1} \log(\varepsilon^3 n)\rceil$ and $s = \frac{1}{8}\varepsilon^{-2} \log(\varepsilon^3 n)$. Then for any sufficiently large n,*

$$\mathbf{P}(A_{r,s}(\mathcal{T})) \geq \varepsilon(\varepsilon^3 n)^{-2/3} .$$

Proof. We first give a lower bound on the probability that $|\mathcal{T}| \geq s$. By (5.3.8), we have $\mathbf{P}(|\mathcal{T}| \geq s) \geq c \sum_{k \geq s} k^{-3/2}(\mu e^{1-\mu})^k$ for some absolute $c > 0$. Recalling that $\mu = 1 - \varepsilon + \frac{2}{3}\varepsilon^2 + O(\varepsilon^3)$, we have that for n large enough,

$$\mu e^{1-\mu} \geq e^{-(\varepsilon+\varepsilon^2)}e^{\varepsilon-\varepsilon^2} \geq e^{-2\varepsilon^2} .$$

Therefore, for $s = \frac{1}{8}\varepsilon^{-2} \log(\varepsilon^3 n)$ this gives that

$$\mathbf{P}(|\mathcal{T}| \geq s) \geq c \sum_{s \leq k \leq 2s} k^{-3/2}e^{-2\varepsilon^2 k} \geq cs(2s)^{-3/2}e^{-4\varepsilon^2 s} \geq \varepsilon(\varepsilon^3 n)^{-1/2+o(1)} .$$

Combining this with the fact that $\{\mathcal{T}_v : v \in L_r\}$ are i.i.d. $\mathrm{PGW}(\mu)$-trees given L_r, we get

$$\mathbf{P}\left(A_{r,s}(\mathcal{T}) \,\middle|\, L_r\right) = 1 - (1 - \mathbf{P}(|\mathcal{T}| \geq s))^{|L_r|} \geq 1 - (1 - \varepsilon(\varepsilon^3 n)^{-1/2+o(1)})^{|L_r|} .$$

Taking expectation over L_r, we conclude that

$$\begin{aligned}
\mathbf{P}\left(A_{r,s}(\mathcal{T})\right) &\geq 1 - \mathbf{E}\left((1 - \varepsilon(\varepsilon^3 n)^{-1/2+o(1)})^{|L_r|}\right) \\
&\geq \varepsilon(\varepsilon^3 n)^{-1/2+o(1)}\mathbf{E}|L_r| - \varepsilon^2(\varepsilon^3 n)^{-1+o(1)}\mathbf{E}|L_r|^2 .
\end{aligned} \tag{5.3.9}$$

For $r = \lceil \frac{1}{8}\varepsilon^{-1} \log(\varepsilon^3 n)\rceil$ we have

$$\mathbf{E}(|L_r|) = \mu^r \geq e^{-(\varepsilon+O(\varepsilon^2))r} \geq (\varepsilon^3 n)^{-1/8+o(1)} ,$$

and by (5.3.4),

$$\mathrm{Var}\,|L_r| = \mu^r \frac{1 - \mu^r}{1 - \mu} \leq e^{-\varepsilon r}2\varepsilon^{-1} \leq 2\varepsilon^{-1}(\varepsilon^3 n)^{-1/8} .$$

Plugging these estimates into (5.3.9), we obtain that

$$\mathbf{P}\left(A_{r,s}(\mathcal{T})\right) \geq \varepsilon(\varepsilon^3 n)^{-5/8+o(1)} \geq \varepsilon(\varepsilon^3 n)^{-2/3},$$

where the last inequality holds for large enough n, as required. ∎

5.3.2 Proof of Theorem 1: Upper bound on the mixing time

By Theorem 5.1.1, it suffices to consider $\tilde{\mathcal{C}}_1$ instead of $\mathcal{C}_1$. As in the previous section, we abbreviate $\tilde{\mathcal{C}}_1^{(2)}$ by $\mathcal{H}$.

For each vertex v in the 2-core $\mathcal{H}$, let $\mathcal{T}_v$ be the PGW-tree attached to v in $\tilde{\mathcal{C}}_1$. Let (S_t) be the lazy random walk on $\tilde{\mathcal{C}}_1$, define $\xi_0 = 0$ and for $j \geq 0$,

$$\xi_{j+1} = \begin{cases} \xi_j + 1 & \text{if } S_{\xi_j+1} = S_{\xi_j}, \\ \min\left\{t > \xi_j : S_t \in \mathcal{H}, \ S_t \neq S_{\xi_j}\right\} & \text{otherwise.} \end{cases}$$

Defining $W_j := S_{\xi_j}$, we observe that (W_j) is a lazy random walk on $\mathcal{H}$. Furthermore, started from any $w \in \mathcal{H}$, there are two options:

(i) Do a step in the 2-core (either stay in w via the lazy rule, which has probability $\frac{1}{2}$, or jump to one of the neighbor of w in $\mathcal{H}$, an event that has probability $d_{\mathcal{H}}(w)/2d_{\tilde{\mathcal{C}}_1}(w)$).

(ii) Enter the PGW-tree attached to w (this happens with probability $d_{\mathcal{T}_w}(w)/2d_{\tilde{\mathcal{C}}_1}(w)$).

It is the latter case that incurs a delay for the random walk on $\tilde{\mathcal{C}}_1$. Since the expected return time to w once entering the tree $\mathcal{T}_w$ is $2(|\mathcal{T}_w| - 1)/d_{\mathcal{T}_w}(w)$, and as the number of excursions to the tree follows a geometric distribution with success probability $1 - d_{\mathcal{T}_w}(w)/2d_{\tilde{\mathcal{C}}_1}(w)$, we infer that

$$\mathbf{E}_w \xi_1 = 1 + \frac{2(|\mathcal{T}_w| - 1)}{d_{\mathcal{T}_w}(w)} \cdot \frac{2d_{\tilde{\mathcal{C}}_1}(w)}{2d_{\tilde{\mathcal{C}}_1}(w) - d_{\mathcal{T}_w}(w)} \leq 4|\mathcal{T}_w|.$$

For some constant $C_1 > 0$ to be specified later, let

$$\ell = C_1 \varepsilon^{-2} \log^2(\varepsilon^3 n), \quad \text{and} \quad a_{v,w}(\ell) = \sum_{j=0}^{\ell-1} \mathbf{P}_v(W_j = w). \tag{5.3.10}$$

It follows that

$$\mathbf{E}_v(\xi_\ell) = \sum_{j=0}^{\ell-1} \sum_{w \in \mathcal{H}} \mathbf{P}_v(S_{\xi_j} = w)\mathbf{E}_w \xi_1$$

$$= \sum_{w \in \mathcal{H}} \sum_{j=0}^{\ell-1} \mathbf{P}_v(W_j = w)\mathbf{E}_w \xi_1 \leq 4 \sum_{w \in \mathcal{H}} a_{v,w}(\ell)|\mathcal{T}_w| \,. \tag{5.3.11}$$

We now wish to bound the last expression via Lemma 5.3.1. Let $v \in \mathcal{K}$. Note that, by definition,

$$\sum_{w \in \mathcal{H}} a_{v,w}(\ell) = \ell = C_1 \varepsilon^{-2} \log^2(\varepsilon^3 n) \,.$$

Moreover, by Proposition 5.2.7, there exists some constant $C_2 > 0$ (which depends on C_1) such that w.h.p.

$$\max_{w \in \mathcal{H}} a_{v,w}(\ell) \leq C_2 \varepsilon^{-1} \log(\varepsilon^3 n) \,.$$

Hence, Lemma 5.3.1 (applied on the sequence $\{a_{v,w}(\ell) : w \in \mathcal{H}\}$) gives that there exists some constant $C > 0$ (depending only on C_1, C_2) such that

$$\sum_{w \in \mathcal{H}} a_{v,w}(\ell)|\mathcal{T}_v| \leq C\varepsilon^{-3} \log^2(\varepsilon^3 n) \quad \text{except with probability } (\varepsilon^3 n)^{-2} \,.$$

Since $|\mathcal{K}| = (\frac{4}{3} + o(1))\varepsilon^3 n$ w.h.p., taking a union bound over the vertices of the kernel while recalling (5.3.11) implies that w.h.p.,

$$\mathbf{E}_v(\xi_\ell) \leq C\varepsilon^{-3} \log^2(\varepsilon^3 n) \quad \text{for all } v \in \mathcal{K} \,. \tag{5.3.12}$$

We next wish to bound the hitting time to the kernel $\mathcal{K}$, defined next:

$$\tau_\mathcal{K} = \min\{t : S_t \in \mathcal{K}\} \,.$$

Define τ_x and τ_S analogously as the hitting times of S_t to the vertex x and the subset S respectively. Recall that from any $v \in \tilde{\mathcal{C}}_1$, after time ξ_1 we will have hit a vertex in the 2-core, hence for any $v \in \tilde{\mathcal{C}}_1$ we have

$$\mathbf{E}_v \tau_\mathcal{K} \leq \mathbf{E}_v \tau_\mathcal{H} + \max_{w \in \mathcal{H}} \mathbf{E}_w \tau_\mathcal{K} \,. \tag{5.3.13}$$

To bound the first summand, since

$$\max_{v \in \tilde{\mathcal{C}}_1} \mathbf{E}_v \tau_{\mathcal{H}} = \max_{w \in \mathcal{H}} \max_{v \in \mathcal{T}_w} \mathbf{E}_v \tau_w \,,$$

it clearly suffices to bound $\mathbf{E}_v \tau_w$ for all $w \in \mathcal{H}$ and $v \in \mathcal{T}_w$. To this end, let $w \in \mathcal{H}$, and let $\tilde{S}_t$ be the lazy random walk on $\mathcal{T}_w$. As usual, define $\tilde{\tau}_v = \min\{t : \tilde{S}_t = v\}$. Clearly, for all $v \in \mathcal{T}_w$ we have $\mathbf{E}_v \tau_w = \mathbf{E}_v \tilde{\tau}_w$. We bound $\mathbf{E}_v \tilde{\tau}_w$ by $\mathbf{E}_v \tilde{\tau}_w + \mathbf{E}_w \tilde{\tau}_v$, i.e., the commute time between v and w. Denote by $\mathrm{R}_{\mathrm{eff}}(v, w)$ the effective resistance between v and w when each edge has unit resistance. The commute time identity of [16] (see also [83]) yields that

$$\mathbf{E}_v \tilde{\tau}_w + \mathbf{E}_w \tilde{\tau}_v \leq 4|\mathcal{T}_w| R_{\mathrm{eff}(v \leftrightarrow w)} \leq 4|\mathcal{T}_w| \operatorname{diam}(\mathcal{T}_w) \,, \tag{5.3.14}$$

Now, Lemmas 5.3.4 and 5.3.5 give that for any $w \in \mathcal{H}$, with probability at least $1 - O(\varepsilon(\varepsilon^3 n)^{-2})$,

$$|\mathcal{T}_w| \leq 6\varepsilon^{-2} \log(\varepsilon^3 n) \quad \text{and} \quad \operatorname{diam}(\mathcal{T}_w) \leq 2\varepsilon^{-1} \log(\varepsilon^3 n) \,. \tag{5.3.15}$$

Since w.h.p. $|\mathcal{H}| = (2 + o(1))\varepsilon^2 n$, we can sum the above over the vertices of $\mathcal{H}$ and conclude that w.h.p., (5.3.15) holds simultaneously for all $w \in \mathcal{H}$. Plugging this in (5.3.14), we deduce that

$$\mathbf{E}_v \tilde{\tau}_w + \mathbf{E}_w \tilde{\tau}_v \leq 48\varepsilon^{-3} \log^2(\varepsilon^3 n) \,,$$

and altogether, as the above holds for every $w \in \mathcal{H}$,

$$\max_{v \in \tilde{\mathcal{C}}_1} \mathbf{E}_v \tau_{\mathcal{H}} \leq 48\varepsilon^{-3} \log^2(\varepsilon^3 n) \,. \tag{5.3.16}$$

For the second summand in (5.3.13), consider $e \in \mathcal{K}$ and let $\mathcal{P}_e$ be the 2-path corresponding to e in the 2-core $\mathcal{H}$. Recall that w.h.p. the longest such 2-path in the 2-core has length $(1 + o(1))\varepsilon^{-1} \log(\varepsilon^3 n)$. Since from each point $v \in \mathcal{P}_e$ we have probability at least $2/|\mathcal{P}_e|$ to hit one of the endpoints of the 2-path (belonging to $\mathcal{K}$) before returning to v, it follows that w.h.p., for every $e \in \mathcal{K}$ and $v \in \mathcal{P}_e$ we have

$$\max_{w \in \mathcal{P}_e} \mathbf{E}_w \#\{t \leq \tau_{\mathcal{K}} : W_t = v\} \leq (\tfrac{1}{2} + o(1))\varepsilon^{-1} \log(\varepsilon^3 n) \,. \tag{5.3.17}$$

We now wish to apply Lemma 5.3.1 to the sequence $a_v = \max_{w \in \mathcal{P}_e} \mathbf{E}_w \#\{t \leq \tau_{\mathcal{K}} : W_t = v\}$.

Since this sequence satisfies

$$\max_{v \in \mathcal{P}_e} a_v \leq (\tfrac{1}{2} + o(1))\varepsilon^{-1} \log(\varepsilon^3 n) , \quad \sum_{v \in \mathcal{P}_e} a_v \qquad \leq (\tfrac{1}{2} + o(1))\varepsilon^{-2} \log^2(\varepsilon^3 n) ,$$

we deduce that there exists some absolute constant $C' > 0$ such that, except with probability $O((\varepsilon^3 n)^{-2})$, every $w \in \mathcal{P}_e$ satisfies

$$\mathbf{E}_w \tau_{\mathcal{K}} \leq C' \varepsilon^{-3} \log^2 \varepsilon^3 n . \tag{5.3.18}$$

Recalling that $e(\mathcal{K}) = (2 + o(1))\varepsilon^3 n$ w.h.p., we deduce that w.h.p. this statement holds simultaneously for all $w \in \mathcal{H}$. Plugging (5.3.16) and (5.3.18) into (5.3.13) we conclude that w.h.p.

$$\mathbf{E}_v \tau_{\mathcal{K}} \leq (C' + 48)\varepsilon^{-3} \log^2 \varepsilon^3 n \text{ for all } v \in \tilde{\mathcal{C}}_1 .$$

Finally, we will now translate these hitting time bounds into an upper bound on the approximate forget time for S_t. Let $\pi_{\mathcal{H}}$ denote the stationary measure on the walk restricted to $\mathcal{H}$:

$$\pi_{\mathcal{H}}(w) = d_{\mathcal{H}}(w)/2e(\mathcal{H}) \quad \text{for } w \in \mathcal{H} .$$

Theorem 5.2.1 enables us to choose some absolute constant $C_1 > 0$ so that ℓ, defined in (5.3.10) as $C_1 \varepsilon^{-2} \log^2(\varepsilon^3 n)$, would w.h.p. satisfy

$$\max_{w \in \mathcal{H}} \left\| \frac{1}{\ell} \sum_{j=1}^{\ell} \mathbf{P}_w(W_j \in \cdot) - \pi_{\mathcal{H}} \right\|_{\mathrm{TV}} \leq \frac{1}{4} . \tag{5.3.19}$$

Define $\bar{\xi}_0 = \tau_{\mathcal{K}}$ and for $j \geq 0$, define $\bar{\xi}_{j+1}$ as we did for ξ_j's, that is,

$$\bar{\xi}_{j+1} = \begin{cases} \bar{\xi}_j + 1 & \text{if } S_{\bar{\xi}_j + 1} = S_{\bar{\xi}_j} , \\ \min\left\{ t > \bar{\xi}_j \ : \ S_t \in \mathcal{H} , \ S_t \neq S_{\bar{\xi}_j} \right\} & \text{otherwise.} \end{cases} .$$

Let Γ be the stopping rule that selects $j \in \{0, \ldots, \ell - 1\}$ uniformly and then stops at $\bar{\xi}_j$. By (5.3.19), w.h.p.

$$\max_{v \in \tilde{\mathcal{C}}_1} \left\| \mathbf{P}_v(S_\Gamma \in \cdot) - \pi_{\mathcal{H}} \right\|_{\mathrm{TV}} \leq \frac{1}{4} .$$

Going back to the definition of the approximate forget time in (5.1.1), taking $\varphi = \pi_{\mathcal{H}}$ with the stopping rule Γ yields $\mathcal{F}_{1/4} \leq \max_{v \in \tilde{\mathcal{C}}_1} \mathbf{E}\Gamma \leq \max_{v \in \tilde{\mathcal{C}}_1} \bar{\xi}_\ell .$

Furthermore, combining (5.3.12) and (5.3.18), we get that w.h.p. for any $v \in \tilde{\mathcal{C}}_1$:

$$\mathbf{E}_v \bar{\xi}_\ell \leq (C + C' + 48)\varepsilon^{-3} \log^2(\varepsilon^3 n) \,.$$

Altogether, we can conclude that the approximate forget time for S_t w.h.p. satisfies that

$$\mathcal{F}_{1/4} \leq \max_{v \in \tilde{\mathcal{C}}_1} \mathbf{E}_v \bar{\xi}_\ell \leq (C + C' + 48)\varepsilon^{-3} \log^2(\varepsilon^3 n) \,.$$

This translates into the required upper bound on t_{MIX} via an application of Theorems 5.1.3 and 5.1.4. ∎

5.3.3 Proof of Theorem 1: Lower bound on the mixing time

As before, by Theorem 5.1.1 it suffices to prove the analogous statement for $\tilde{\mathcal{C}}_1$.

Let r, s be as in Lemma 5.3.6, i.e.,

$$r = \lceil \tfrac{1}{8}\varepsilon^{-1} \log(\varepsilon^3 n) \rceil \quad \text{and} \quad s = \tfrac{1}{8}\varepsilon^{-2} \log(\varepsilon^3 n) \,.$$

Let $\mathcal{T}_v$ for $v \in \mathcal{H}$ be the $\text{PGW}(\mu)$-tree that is attached to the vertex v. Lemma 5.3.6 gives that when n is sufficiently large, every $v \in \mathcal{H}$ satisfies

$$\mathbf{P}(A_{r,s}(\mathcal{T}_v)) \geq \varepsilon(\varepsilon^3 n)^{-2/3} \,.$$

Since $|\mathcal{H}| = (2 + o(1))\varepsilon^2 n$ w.h.p. (recall Theorem 5.1.1), and since $\{\mathcal{T}_v : v \in \mathcal{H}\}$ are i.i.d. given $\mathcal{H}$, we can conclude that w.h.p. there exists some $\rho \in \mathcal{H}$ such that $A_{r,s}(\mathcal{T}_\rho)$ holds. Let $\rho \in \mathcal{H}$ therefore be such a vertex.

Let (S_t) be a lazy random walk on $\tilde{\mathcal{C}}_1$ and π be its stationary distribution. As usual, let $\tau_v := \min\{t : S_t = v\}$. We wish to prove that

$$\max_{w \in \mathcal{T}_\rho} \mathbf{P}_w \left(\tau_\rho \geq \tfrac{2}{3}rs \right) \geq \tfrac{1}{3} \,. \tag{5.3.20}$$

For $w \in \mathcal{T}_\rho$, let $\mathcal{T}_w$ be the entire subtree rooted at w. Further let L_r be the vertices of the r-th level of $\mathcal{T}_\rho$. By our assumption on $\mathcal{T}_\rho$, there is some $\xi \in L_r$ such that $|\mathcal{T}_\xi| \geq s$.

We will derive a lower bound on $\mathbf{E}_\xi \tau_\rho$ from the following well-known connection between hitting-times of random walks and flows on electrical networks (see [83] and also [63, Proposition 2.19]).

Lemma 5.3.7 ([83]). *Given a graph $G = (V, E)$ with a vertex z and a subset of vertices Z not containing z, let $v(\cdot)$ be the voltage when a unit current flows from z to Z and the*

voltage is 0 on Z. Then $\mathbf{E}_z \tau_Z = \sum_{x \in V} d(x) v(x)$.

In our setting, we consider the graph $\tilde{\mathcal{C}}_1$. Clearly, the effective resistance between ρ and ξ satisfies $R_{\mathrm{eff}}(\rho \leftrightarrow \xi) = r$. If a unit current flows from ξ to ρ and $v(\rho) = 0$, it follows from Ohm's law that $v(\xi) = r$. Notice that for any $w \in \mathcal{T}_\xi$, the flow between w and ξ is 0. Altogether, we deduce that

$$v(w) = r \quad \text{for all } w \in \mathcal{T}_\xi.$$

Therefore, Lemma 5.3.7 implies that

$$\mathbf{E}_\xi \tau_\rho \geq r|\mathcal{T}_\xi| \geq rs.$$

Clearly, if $w^\star \in \mathcal{T}_\rho$ attains $\max\{\mathbf{E}_w \tau_\rho : w \in \mathcal{T}_\rho\}$ then clearly

$$\mathbf{E}_{w^\star} \tau_\rho \leq \tfrac{2}{3}rs + \mathbf{P}_{w^\star}\left(\tau_\rho \geq \tfrac{2}{3}rs\right) \mathbf{E}_{w^\star} \tau_\rho.$$

On the other hand,

$$\mathbf{E}_{w^\star} \tau_\rho \geq \mathbf{E}_\xi \tau_\rho \geq rs,$$

hence we obtain (5.3.20).

Recall that w.h.p. $|\tilde{\mathcal{C}}_1| = (2+o(1))\varepsilon n$. Together with Lemma 5.3.5, we deduce that w.h.p. every $v \in \mathcal{H}$ satisfies

$$|\mathcal{T}_v| \leq 6\varepsilon^{-2} \log(\varepsilon^3 n) = o(|\tilde{\mathcal{C}}_1|).$$

In particular, $|\mathcal{T}_\rho| = o(|\tilde{\mathcal{C}}_1|)$, and so (as it is a tree) $\pi(\mathcal{T}_\rho) = o(1)$. However, (5.3.20) states that with probability at least $\frac{1}{3}$, the random walk started at some $w \in \mathcal{T}_\rho$ does not escape from $\mathcal{T}_\rho$, hence

$$\max_{w \in \tilde{\mathcal{C}}_1} \|\mathbf{P}_w(S_{2rs/3} \in \cdot) - \pi\|_{\mathrm{TV}} \geq \tfrac{1}{4}.$$

where π is the stationary measure for the random walk S_t on $\tilde{\mathcal{C}}_1$. In other words, we have that

$$t_{\mathrm{MIX}}\left(\tfrac{1}{4}\right) \geq \tfrac{2}{3}rs = \tfrac{1}{96}\varepsilon^{-3} \log^2(\varepsilon^3 n),$$

as required. ∎

5.4 Mixing in the subcritical regime

In this section, we give the proof of Theorem 2. By Theorem 1 and the well known duality between the subcritical and supercritical regimes (see [60]), it suffices to establish the statement for the subcritical regime of $\mathcal{G}(n, p)$.

For the upper bound, by results of [11] and [60] (see also [69]), we know that the largest component has size $O(\varepsilon^{-2}\log(\varepsilon^3 n))$ w.h.p., and by results of [61], the largest diameter of a component is w.h.p. $O(\varepsilon^{-1}\log(\varepsilon^3 n))$. Therefore, by the commute time identity (5.3.14) the maximal hitting time to a vertex is $O(\varepsilon^{-3}\log^2(\varepsilon^3 n))$ uniformly for all components, and using the well-known fact that $t_{\text{MIX}} = O(\max_{x,y}\mathbf{E}_x\tau_y)$ (see, e.g., [3, Chapter 2]) we arrive at the desired upper bound on the mixing time.

In order to establish the lower bound, we will demonstrate the existence of a component with a certain structure, and show that the order of the mixing time on this particular component matches the above upper bound.

To find this component, we apply the usual exploration process until εn vertices are exposed. By definition, each component revealed is a Galton-Watson tree (the exploration process does not expose the tree-excess) where the offspring distribution is stochastically dominated by $\text{Bin}(n, \frac{1-\varepsilon}{n})$ and stochastically dominates $\text{Bin}(n, \frac{1-2\varepsilon}{n})$.

It is well known (see, e.g., [55, equation (1.12)]) that for any $\lambda > 0$,

$$\left\|\text{Bin}(n, \tfrac{\lambda}{n}) - \text{Po}(\lambda)\right\|_{\text{TV}} \le \lambda^2/n\,.$$

It follows that when discovering the first εn vertices, we can approximate the binomial variables by Poisson variables, at the cost of a total error of at most $\varepsilon n(1/n) = \varepsilon = o(1)$.

Lemma 5.4.1. *With high probability, once εn vertices are exposed in the exploration process, we will have discovered at least $\varepsilon^2 n/2$ components.*

Proof. Notice that each discovered component is stochastically dominated (with respect to containment) by a Poisson$(1-\varepsilon)$-Galton-Watson tree. Thus, the probability that the first $\varepsilon^2 n/2$ components contain more than εn vertices is bounded by the probability that the total size of $\varepsilon^2 n/2$ independent $\text{PGW}(1-\varepsilon)$-trees is larger than εn. The latter can be estimated (using Chebyshev's inequality and Claim 5.3.2) by

$$\mathbf{P}\Big(\sum_{i=1}^{\varepsilon^2 n/2}|\mathcal{T}_i| \ge \varepsilon n\Big) \le \frac{\varepsilon^2 n\varepsilon^{-3}}{(\varepsilon n/2)^2} = 4(\varepsilon^3 n)^{-1} = o(1)\,. \qquad \blacksquare$$

For a rooted tree $\mathcal{T}$, we define the following event, analogous to the event $A_{r,s}(\mathcal{T})$ from Subsection 5.3.1:

$$B_{r,s}(\mathcal{T}) := \{\exists v, w \in \mathcal{T} \text{ such that } |\mathcal{T}_v| \ge s\,, |\mathcal{T}_w| \ge s \text{ and } \text{dist}(v, w) = r\}\,.$$

The next lemma estimates the probability that the above defined event occurs in a PGW-tree.

Lemma 5.4.2. *Let $\mathcal{T}$ be a* PGW$(1 - 2\varepsilon)$*-tree and set* $r = \lceil \frac{1}{20}\varepsilon^{-1}\log(\varepsilon^3 n)\rceil$ *and* $s = \frac{1}{64}\varepsilon^{-2}\log(\varepsilon^3 n)$. *Then for some $c > 0$ and any sufficiently large n,*

$$\mathbf{P}(B_{r,s}(\mathcal{T})) \geq c\varepsilon(\varepsilon^3 n)^{-1/2}.$$

Proof. The proof follows the general argument of Lemma 5.3.6. By Lemma 5.3.4,

$$\mathbf{P}(L_{1/\varepsilon} \neq \emptyset) \asymp \varepsilon.$$

Combined with the proof of Claim 5.3.2 (see (5.3.4) in particular), we get that

$$\mathbf{E}\big(|L_{1/\varepsilon}| \mid L_{1/\varepsilon} \neq \emptyset\big) \asymp \varepsilon^{-1}, \text{ and } \mathrm{Var}\big(|L_{1/\varepsilon}| \mid L_{1/\varepsilon} \neq \emptyset\big) \asymp \varepsilon^{-2}.$$

Applying Chebyshev's inequality, we get that for some constants $c_1, c_2 > 0$

$$\mathbf{P}\left(|L_{1/\varepsilon}| > c_1\varepsilon^{-1} \mid L_{1/\varepsilon} \neq \emptyset\right) \geq c_2.$$

Repeating the arguments for the proof of Lemma 5.3.6, we conclude that for a PGW$(1-2\varepsilon)$-tree $\mathcal{T}$, the probability that the event $A_{r,s}(\mathcal{T})$ occurs (using r, s as defined in the current lemma) is at least $\varepsilon(\varepsilon^3 n)^{-1/4}$ for n large enough. Thus (by the independence of the subtrees rooted in the $(1/\varepsilon)$-th level),

$$\mathbf{P}\left(\bigcup\left\{A_{r,s}(\mathcal{T}_u) \cap A_{r,s}(\mathcal{T}_{u'}) : \begin{smallmatrix} u,u' \in L_{1/\varepsilon} \\ u \neq u' \end{smallmatrix}\right\} \,\Big|\, |L_{1/\varepsilon}| > c_1\varepsilon^{-1}\right) \geq c(\varepsilon^3 n)^{-1/2}$$

for some $c > 0$. Altogether, we conclude that for some $c' > 0$,

$$\mathbf{P}\left(\bigcup\left\{A_{r,s}(\mathcal{T}_u) \cap A_{r,s}(\mathcal{T}_{u'}) : \begin{smallmatrix} u,u' \in L_{1/\varepsilon} \\ u \neq u' \end{smallmatrix}\right\}\right) \geq c'\varepsilon(\varepsilon^3 n)^{-1/2},$$

which immediately implies that required bound on $\mathbf{P}(B_{r,s}(\mathcal{T}))$. ∎

Combining Lemmas 5.4.1 and 5.4.2, we conclude that w.h.p., during our exploration process we will find a tree $\mathcal{T}$ which satisfies the event $B_{r,s}(\mathcal{T})$ for r, s as defined in Lemma 5.4.2. Next, we will show that the component of $\mathcal{T}$ is indeed a tree, namely, it has no tree-excess. Clearly, edges belonging to the tree-excess can only appear between vertices that belong either to the same level or to successive levels (the root of the tree $\mathcal{T}$ is defined to be the vertex in $\mathcal{T}$ that is first exposed). Therefore, the total number of candidates for such edges can be bounded by $4\sum_i |L_i|^2$ where L_i is the i-th level of vertices in the tree. The next claim provides an upper bound for this sum.

Claim 5.4.3. *Let r, s be defined as in Lemma 5.4.2. Then the $\mathrm{PGW}(1-\varepsilon)$-tree $\mathcal{T}$ satisfies* $\mathbf{E}\big[\sum_i |L_i|^2 \mid B_{r,s}(\mathcal{T})\big] = O\big(\varepsilon^{-3}\sqrt{\varepsilon^3 n}\big)$.

Proof. Recalling Claim 5.3.2 and in particular equation (5.3.4), it follows that $\mathbf{E}\big(\sum_i |L_i|^2\big) \leq \varepsilon^{-2}$. Lemma 5.4.2 now implies the required upper bound. $\blacksquare$

By the above claim and Markov's inequality, we deduce that w.h.p. there are, say, $O\big(\varepsilon^{-3}(\varepsilon^3 n)^{2/3}\big)$ candidates for edges in the tree-excess of the component of $\mathcal{T}$. Crucially, whether or not these edges appear is independent of the exploration process, hence the probability that any of them appears is at most $O\big((\varepsilon^3 n)^{-1/3}\big) = o(1)$. Altogether, we may assume that the component of $\mathcal{T}$ is indeed a tree which satisfies the event $B_{r,s}(\mathcal{T})$.

It remains to establish the lower bound on the mixing time of the random walk on the tree $\mathcal{T}$. Let v, w be two distinct vertices in the r-th level satisfying $|\mathcal{T}_v| \geq s$ and $|\mathcal{T}_w| \geq s$. By the same arguments used to prove (5.3.20), we have that

$$\max_{u \in \mathcal{T}_v} \mathbf{P}_u(\tau_w \geq 10^{-3} rs) \geq 1 - 10^{-3}\,.$$

Recall that w.h.p. $|\mathcal{T}| \leq 6\varepsilon^{-2}\log(\varepsilon^3 n) = 384s$. It now follows that w.h.p. the mixing time of the random walk on this components satisfies

$$t_{\mathrm{MIX}}(\delta) \geq 10^{-3} rs\,, \ \text{ for } \delta = \tfrac{1}{384} - 10^{-3} \geq 10^{-3}\,.$$

The lower bound on $t_{\mathrm{MIX}}(\tfrac{1}{4})$ now follows from the definition of r, s. $\blacksquare$

Bibliography

[1] M. Aizenman and R. Holley, *Rapid convergence to equilibrium of stochastic Ising models in the Dobrushin Shlosman regime*, (Minneapolis, Minn., 1984), IMA Vol. Math. Appl., vol. 8, Springer, New York, 1987, pp. 1–11.

[2] D. Aldous, *Random walks on finite groups and rapidly mixing Markov chains*, Seminar on probability, XVII, Lecture Notes in Math., vol. 986, Springer, Berlin, 1983, pp. 243–297.

[3] D. Aldous and J. A. Fill, *Reversible Markov Chains and Random Walks on Graphs*. In preparation, `http://www.stat.berkeley.edu/~aldous/RWG/book.html`.

[4] D. Aldous, L. Lovász, and P. Winkler, *Mixing times for uniformly ergodic Markov chains*, Stochastic Process. Appl. **71** (1997), no. 2, 165–185.

[5] R. J. Baxter, *Exactly solved models in statistical mechanics*, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London, 1989. Reprint of the 1982 original.

[6] I. Benjamini, G. Kozma, and N. C. Wormald, *The mixing time of the giant component of a random graph*. preprint, available at `http://arxiv.org/abs/math/0610459`.

[7] I. Benjamini and E. Mossel, *On the mixing time of a simple random walk on the super critical percolation cluster*, Probab. Theory Related Fields **125** (2003), no. 3, 408–420.

[8] N. Berger, C. Kenyon, E. Mossel, and Y. Peres, *Glauber dynamics on trees and hyperbolic graphs*, Probab. Theory Related Fields **131** (2005), no. 3, 311–340.

[9] P. M. Bleher, J. Ruiz, and V. A. Zagrebnov, *On the purity of the limiting Gibbs state for the Ising model on the Bethe lattice*, J. Statist. Phys. **79** (1995), no. 1-2, 473–482.

[10] B. Bollobás, *Random graphs*, 2nd ed., Cambridge Studies in Advanced Mathematics, vol. 73, Cambridge University Press, Cambridge, 2001.

[11] B. Bollobás, *The evolution of random graphs*, Trans. Amer. Math. Soc. **286** (1984), no. 1, 257–274.

[12] R. Bubley and M. Dyer, *Path coupling: A technique for proving rapid mixing in Markov chains*, Proceedings of the 38th Annual Symposium on Foundations of Computer Science (FOCS), 1997, pp. 223–231.

[13] J. M. Carlson, J. T. Chayes, L. Chayes, J. P. Sethna, and D. J. Thouless, *Bethe lattice spin glass: the effects of a ferromagnetic bias and external fields. I. Bifurcation analysis*, J. Statist. Phys. **61** (1990), no. 5-6, 987–1067.

[14] J. M. Carlson, J. T. Chayes, L. Chayes, J. P. Sethna, and D. J. Thouless, *Critical Behavior of the Bethe Lattice Spin Glass*, Europhys. Lett. **106** (1988), no. 5, 355–360.

[15] J. M. Carlson, J. T. Chayes, J. P. Sethna, and D. J. Thouless, *Bethe lattice spin glass: the effects of a ferromagnetic bias and external fields. II. Magnetized spin-glass phase and the de Almeida-Thouless line*, J. Statist. Phys. **61** (1990), no. 5-6, 1069–1084.

[16] A. K. Chandra, P. Raghavan, W. L. Ruzzo, R. Smolensky, and P. Tiwari, *The electrical resistance of a graph captures its commute and cover times*, Comput. Complexity **6** (1996/97), no. 4, 312–340.

[17] M.-F. Chen, *Trilogy of couplings and general formulas for lower bound of spectral gap*, Probability towards 2000 (New York, 1995), Lecture Notes in Statist., vol. 128, Springer, New York, 1998, pp. 123–136.

[18] P. Diaconis and L. Saloff-Coste, *Comparison techniques for random walk on finite groups*, Ann. Probab. **21** (1993), no. 4, 2131–2156.

[19] P. Diaconis and L. Saloff-Coste, *Comparison theorems for reversible Markov chains*, Ann. Appl. Probab. **3** (1993), no. 3, 696–730.

[20] P. Diaconis and L. Saloff-Coste, *Logarithmic Sobolev inequalities for finite Markov chains*, Ann. Appl. Probab. **6** (1996), no. 3, 695–750.

[21] P. Diaconis and L. Saloff-Coste, *Nash inequalities for finite Markov chains*, J. Theoret. Probab. **9** (1996), no. 2, 459–510.

[22] J. Ding, J. H. Kim, E. Lubetzky, and Y. Peres, *Anatomy of a young giant component in the random graph*, Random Structures Algorithms, to appear.

[23] J. Ding, J. H. Kim, E. Lubetzky, and Y. Peres, *Diameters in supercritical random graphs via first passage percolation*, Combin. Probab. Comput. **19** (2010), no. 5-6, 729–751.

[24] J. Ding, E. Lubetzky, and Y. Peres, *Mixing time of critical Ising model on trees is polynomial in the height*, Comm. Math. Phys. **295** (2010), no. 1, 161–207, DOI 10.1007/s00220-009-0978-y. MR **2585995 (2011b:**82025)

[25] J. Ding, E. Lubetzky, and Y. Peres, *The mixing time evolution of Glauber dynamics for the mean-field Ising model*, Comm. Math. Phys. **289** (2009), no. 2, 725–764.

[26] J. Ding, E. Lubetzky, and Y. Peres, *Censored Glauber dynamics for the mean field Ising model*, J. Stat. Phys. **137** (2009), no. 3, 407–458.

[27] J. Ding, E. Lubetzky, and Y. Peres, *Total variation cutoff in birth-and-death chains*, Probab. Theory Related Fields **146** (2010), no. 1-2, 61–85.

[28] J. Ding, E. Lubetzky, and Y. Peres, *Mixing times for near-critical random graphs.* to appear in Annals of Probability.

[29] J. Ding and Y. Peres, *Mixing times for the Ising model: a uniform lower bound.* to appear in AIHP.

[30] C. Domb and J. L. Lebowitz (eds.), *Phase transitions and critical phenomena. Vol. 20*, Academic Press, San Diego, CA, 2001. Cumulative author, title and subject index, including tables of contents, Vol. 1–19.

[31] R. S. Ellis, *Entropy, large deviations, and statistical mechanics*, Grundlehren der Mathematischen Wissenschaften, vol. 271, Springer-Verlag, New York, 1985.

[32] R. S. Ellis and J. L. Monroe, *A simple proof of the* GHS *and further inequalities*, Comm. Math. Phys. **41** (1975), 33–38.

[33] R. S. Ellis and C. M. Newman, *Limit theorems for sums of dependent random variables occurring in statistical mechanics*, Z. Wahrsch. Verw. Gebiete **44** (1978), no. 2, 117–139.

[34] R. S. Ellis, C. M. Newman, and J. S. Rosen, *Limit theorems for sums of dependent random variables occurring in statistical mechanics. II. Conditioning, multiple phases, and metastability*, Z. Wahrsch. Verw. Gebiete **51** (1980), no. 2, 153–169.

[35] P. Erdős and A. Rényi, *On random graphs. I*, Publ. Math. Debrecen **6** (1959), 290–297.

[36] W. Evans, C. Kenyon, Y. Peres, and L. J. Schulman, *Broadcasting on trees and the Ising model*, Ann. Appl. Probab. **10** (2000), no. 2, 410–433.

[37] N. Fountoulakis and B. A. Reed, *Faster mixing and small bottlenecks*, Probab. Theory Related Fields **137** (2007), no. 3-4, 475–486.

[38] N. Fountoulakis and B. A. Reed, *The evolution of the mixing rate of a simple random walk on the giant component of a random graph*, Random Structures Algorithms **33** (2008), no. 1, 68–86.

[39] H.-O. Georgii, O. Häggström, and C. Maes, *The random geometry of equilibrium phases*, Phase transitions and critical phenomena, Vol. 18, Phase Transit. Crit. Phenom., vol. 18, Academic Press, San Diego, CA, 2001, pp. 1–142.

[40] R. B. Griffiths, C. A. Hurst, and S. Sherman, *Concavity of magnetization of an Ising ferromagnet in a positive external field*, J. Mathematical Phys. **11** (1970), 790–795.

[41] R. B. Griffiths, C.-Y. Weng, and J. S. Langer, *Relaxation Times for Metastable States in the Mean-Field Model of a Ferromagnet*, Phys. Rev. **149** (1966), 301 –305.

[42] T. P. Hayes and A. Sinclair, *A general lower bound for mixing of single-site dynamics on graphs*, Ann. Appl. Probab. **17** (2007), no. 3, 931–952. Preliminary version appeared in Proceedings of IEEE FOCS 2005, pp. 511-520.

[43] M. Heydenreich and R. van der Hofstad, *Random graph asymptotics on high-dimensional tori II. Volume, diameter and mixing time*, Probab. Theory Related Fields, to appear.

[44] P. C. Hohenberg and B. I. Halperin, *Theory of dynamic critical phenomena*, Rev. Mod. Phys. **49** (1977), no. 3, 435–479.

[45] D. Ioffe, *Extremality of the disordered state for the Ising model on general trees*, Trees (Versailles, 1995), Progr. Probab., vol. 40, Birkhäuser, Basel, 1996, pp. 3–14 (English, with English and French summaries).

[46] D. Ioffe, *On the extremality of the disordered state for the Ising model on the Bethe lattice*, Lett. Math. Phys. **37** (1996), no. 2, 137–143.

[47] M. Jerrum and A. Sinclair, *Approximate counting, uniform generation and rapidly mixing Markov chains*, Inform. and Comput. **82** (1989), no. 1, 93–133.

[48] M. Jerrum, J.-B. Son, P. Tetali, and E. Vigoda, *Elementary bounds on Poincaré and log-Sobolev constants for decomposable Markov chains*, Ann. Appl. Probab. **14** (2004), no. 4, 1741–1765.

[49] J. H. Kim, *Poisson cloning model for random graphs*, International Congress of Mathematicians. Vol. III, Eur. Math. Soc., Zürich, 2006, pp. 873–897.

[50] K. B. Lauritsen and H. C. Fogedby, *Critical exponents from power spectra*, J. Statist. Phys. **72** (1993), no. 1, 189–205.

[51] J. L. Lebowitz, *GHS and other inequalities*, Comm. Math. Phys. **35** (1974), 87–92.

[52] L. Le Cam, *Notes on asymptotic methods in statistical decision theory*, Centre de Recherches Mathématiques, Université de Montréal, Montreal, Que., 1974.

[53] D. A. Levin, M. J. Luczak, and Y. Peres, *Glauber dynamics for the mean-field Ising model: cut-off, critical power law, and metastability*, Probab. Theory Related Fields **146** (2010), no. 1-2, 223–265.

[54] D. A. Levin, Y. Peres, and E. Wilmer, *Markov Chains and Mixing Times*, 2007. In preparation, available at `http://www.uoregon.edu/~dlevin/MARKOV/`.

[55] T. Lindvall, *Lectures on the coupling method*, Dover Publications Inc., Mineola, NY, 2002. Corrected reprint of the 1992 original.

[56] L. Lovász and R. Kannan, *Faster mixing via average conductance*, Annual ACM Symposium on Theory of Computing (Atlanta, GA, 1999), ACM, New York, 1999, pp. 282–287 (electronic).

[57] L. Lovász and P. Winkler, *Mixing of random walks and other diffusions on a graph*, Surveys in combinatorics, 1995 (Stirling), London Math. Soc. Lecture Note Ser., vol. 218, Cambridge Univ. Press, Cambridge, 1995, pp. 119–154.

[58] L. Lovász and P. Winkler, *Mixing times*, Microsurveys in discrete probability (Princeton, NJ, 1997), DIMACS Ser. Discrete Math. Theoret. Comput. Sci., vol. 41, Amer. Math. Soc., Providence, RI, 1998, pp. 85–133.

[59] E. Lubetzky and A. Sly, *Cutoff for the Ising model on the lattice* (2009). Preprint.

[60] T. Łuczak, *Component behavior near the critical point of the random graph process*, Random Structures Algorithms **1** (1990), no. 3, 287–310.

[61] T. Łuczak, *Random trees and random graphs*, Proceedings of the Eighth International Conference "Random Structures and Algorithms" (Poznan, 1997), 1998, pp. 485–500.

[62] R. Lyons, *The Ising model and percolation on trees and tree-like graphs*, Comm. Math. Phys. **125** (1989), no. 2, 337–353.

[63] R. Lyons with Y. Peres, *Probability on Trees and Networks*, Cambridge University Press, 2008. In preparation. Current version is available at `http://mypage.iu.edu/~rdlyons/prbtree/book.pdf`.

[64] F. Martinelli, *Lectures on Glauber dynamics for discrete spin models*, Lectures on probability theory and statistics (Saint-Flour, 1997), Lecture Notes in Math., vol. 1717, Springer, Berlin, 1999, pp. 93–191.

[65] F. Martinelli, *On the two-dimensional dynamical Ising model in the phase coexistence region*, J. Statist. Phys. **76** (1994), no. 5-6, 1179–1246.

[66] F. Martinelli, A. Sinclair, and D. Weitz, *Glauber dynamics on trees: boundary conditions and mixing time*, Comm. Math. Phys. **250** (2004), no. 2, 301–334.

[67] P. Mathieu and E. Remy, *Isoperimetry and heat kernel decay on percolation clusters*, Ann. Probab. **32** (2004), no. 1A, 100–128.

[68] A. Murakami and M. Yamasaki, *Nonlinear potentials on an infinite network*, Mem. Fac. Sci. Shimane Univ. **26** (1992), 15–28.

[69] A. Nachmias and Y. Peres, *Component sizes of the random graph outside the scaling window*, ALEA Lat. Am. J. Probab. Math. Stat. **3** (2007), 133–142 (electronic).

[70] A. Nachmias and Y. Peres, *Critical random graphs: diameter and mixing time*, Ann. Probab. **36** (2008), no. 4, 1267–1286.

[71] Ş. Nacu, *Glauber Dynamics on the Cycle is Monotone*, Probability Theory and Related Fields **127** (2003), 177-185.

[72] C. St. J. A. Nash-Williams, *Random walk and electric currents in networks*, Proc. Cambridge Philos. Soc. **55** (1959), 181–194.

[73] R. Pemantle and Y. Peres, *The critical Ising model on trees, concave recursions and nonlinear capacity* (2005). preprint.

[74] Y. Peres, Lectures on "Mixing for Markov Chains and Spin Systems" (University of British Columbia, August 2005). Summary available at `http://www.stat.berkeley.edu/~peres/ubc.pdf`.

[75] Y. Peres and P. Winkler, *Can extra updates delay mixing?*. In preparation.

[76] J. Pitman, *Enumerations of trees and forests related to branching processes and random walks*, Microsurveys in discrete probability (Princeton, NJ, 1997), DIMACS Ser. Discrete Math. Theoret. Comput. Sci., vol. 41, Amer. Math. Soc., Providence, RI, 1998, pp. 163–180.

[77] C. J. Preston, *Gibbs states on countable sets*, Cambridge University Press, London, 1974. Cambridge Tracts in Mathematics, No. 68.

[78] O. Riordan and N. Wormald, *The diameter of sparse random graphs*, Combin. Probab. Comput. **19** (2010), no. 5-6, 835–926.

[79] L. Saloff-Coste, *Lectures on finite Markov chains*, Lectures on probability theory and statistics (Saint-Flour, 1996), Lecture Notes in Math., vol. 1665, Springer, Berlin, 1997, pp. 301–413.

[80] A. Sinclair, *Algorithms for random generation and counting*, Progress in Theoretical Computer Science, Birkhäuser Boston Inc., Boston, MA, 1993. A Markov chain approach.

[81] P. M. Soardi, *Morphisms and currents in infinite nonlinear resistive networks*, Potential Anal. **2** (1993), no. 4, 315–347.

[82] P. M. Soardi, *Potential theory on infinite networks*, Lecture Notes in Mathematics, vol. 1590, Springer-Verlag, Berlin, 1994.

[83] P. Tetali, *Random walks and the effective resistance of networks*, J. Theoret. Probab. **4** (1991), no. 1, 101–109.

[84] F.-G. Wang and C.-K. Hu, *Universality in dynamic critical phenomena*, Phys. Rev. E **56** (1997), no. 2, 2310–2313.

[85] D. B. Wilson, *Mixing times of Lozenge tiling and card shuffling Markov chains*, Ann. Appl. Probab. **14** (2004), no. 1, 274–325.