Optimal Mixing for Randomly Sampling Edge Colorings on Trees Down to the Max Degree

Charlie Carlson[*]   Xiaoyu Chen[†]   Weiming Feng[‡]   Eric Vigoda[*]

Abstract

We address the convergence rate of Markov chains for randomly generating an edge coloring of a given tree. Our focus is on the Glauber dynamics which updates the color at a randomly chosen edge in each step. For a tree T with n vertices and maximum degree Δ, when the number of colors q satisfies $q \geq \Delta + 2$ then we prove that the Glauber dynamics has an optimal relaxation time of $O(n)$, where the relaxation time is the inverse of the spectral gap. This is optimal in the range of q in terms of Δ as Dyer, Goldberg, and Jerrum (2006) showed that the relaxation time is $\Omega(n^3)$ when $q = \Delta + 1$. For the case $q = \Delta + 1$, we show that an alternative Markov chain which updates a pair of neighboring edges has relaxation time $O(n)$. Moreover, for the Δ-regular complete tree we prove $O(n \log^2 n)$ mixing time bounds for the respective Markov chain. Our proofs establish approximate tensorization of variance via a novel inductive approach, where the base case is a tree of height $\ell = O(\Delta^2 \log^2 \Delta)$, which we analyze using a canonical paths argument.

1 Introduction

We study the Glauber dynamics for sampling q-edge colorings on trees of maximum degree Δ. Given a set of colors $[q] := \{1, 2, \ldots, q\}$, a proper edge coloring $\sigma \in [q]^E$ assigns each edge a color such that adjacent edges receive distinct colors. For a graph $G = (V, E)$, let Ω denote the set of proper q-edge colorings of G, and let μ denote the uniform distribution over Ω. We focus on a given graph G and integer $q \geq 2$, generating a q-coloring almost uniformly at random. More precisely, given G, q and $\varepsilon > 0$, our goal is to generate a coloring $\sigma \in \Omega$ from a probability distribution ν which is within total variation distance $\leq \varepsilon$ of the uniform distribution μ, and to do so in time polynomial in $n = |V|$ and $\log(1/\varepsilon)$.

The Glauber dynamics is a simple Markov chain which we study for this sampling problem. The dynamics operates as follows. From a coloring $X_t \in \Omega$, we choose an edge $e \in E$ and color $c \in [q]$ uniformly at random. If no neighbor of e has color c in X_t then we set $X_{t+1}(e) = c$ and otherwise we set $X_{t+1}(e) = X_t(e)$. For all other edges $f \neq e$ we set $X_{t+1}(f) = X_t(f)$. If the maximum degree of G is Δ then when $q \geq 2\Delta + 2$ then the Glauber dynamics is ergodic and the unique stationary distribution is the uniform distribution μ.

The *mixing time* T_{mix} is the number of steps t, from the worst initial state X_0, so that the distribution of X_t is within total variation distance $\leq 1/4$ of μ. The *relaxation time* T_{rel} is the inverse of the spectral gap. More precisely, let P denote the $N \times N$ transition matrix for the Glauber dynamics where $N = |\Omega|$, and let $\lambda_1 = 1 > \lambda_2 \geq \ldots \lambda_N > -1$ denote the eigenvalues of P. Then, $T_{\mathrm{rel}} := (1 - \lambda_*)^{-1}$ where $\lambda_* = \max\{\lambda_2, |\lambda_N|\}$. We say the relaxation time is optimal if $T_{\mathrm{rel}} = O(n)$ since it is easy to establish the lower bound of $T_{\mathrm{rel}} = \Omega(n)$. The optimal mixing time bound is $O(n \log n)$ [HS07].

For graphs of maximum degree Δ, there always exists a q-edge coloring when $q \geq 2\Delta + 1$, and hence this presents a natural barrier at $q = 2\Delta + 1$ for general graphs. For general graphs of maximum degree Δ when $\Delta = O(1)$ the following results were established. Abdodolazimi, Liu, and Oveis Gharan [ALG21] proved $O(n \log n)$ mixing when $q > (10/3)\Delta$; this improves upon the $q > (11/3 - \varepsilon_0)\Delta$ for a small, fixed $\varepsilon_0 > 0$ which one obtains via vertex q-colorings [CDM+19, Vig00]. A near-optimal bound was obtained by Wang, Zhang and Zhang [WZZ24] who proved $O(n \log n)$ mixing time of Glauber dynamics when $q > (2 + o(1))\Delta$ for any graph of maximum degree Δ for $\Delta = O(1)$.

If we restrict attention to trees then the Glauber dynamics is ergodic for all $q \geq \Delta + 1$, and hence the natural threshold is at $q = \Delta + 1$ for trees. Understanding the convergence rate of the Glauber dynamics on trees is interesting as it can provide intuition for its behavior on locally tree-like graphs such as sparse random graphs. In fact, locally tree-like graphs serve as the crucial gadget for hardness of approximate counting results [Sly10, GŠV15] by utilizing the stationary properties of the Gibbs distribution which are directly related to the mixing properties.

Delcourt, Heinrich, and Perarnau [DHP20] proved poly(n) mixing time for the Glauber dynamics on trees with maximum degree Δ when $q \geq \Delta + 1$. Can we obtain optimal mixing for trees down to $q \geq \Delta + 1$? For a path with $q = 3$ colors (which corresponds to $q = \Delta + 1$ for $\Delta = 2$), Dyer, Goldberg, and Jerrum [DGJ06] proved that the Glauber dynamics has mixing time $\Theta(n^3 \log n)$ and hence relaxation time $T_{\mathrm{rel}} = \Omega(n^3)$. Therefore, optimal mixing and optimal relaxation time does not hold for the Glauber dynamics when $q = \Delta + 1$ for all trees of maximum degree Δ.

We prove that with one additional color optimal relaxation time of the Glauber dynamics holds for all trees of maximum degree Δ.

Theorem 1. *For any tree with maximum degree Δ and n vertices, and any $q \geq \Delta + 2$, then the relaxation time of the Glauber dynamics for q-edge-colorings is $O_{\Delta,q}(n)$, where the big-O notation hides a factor*

depending only on Δ and q.

Moreover, if we consider a slightly more robust dynamics then we can obtain optimal relaxation time down to $\Delta + 1$ colors. We consider the heat-bath dynamics on a pair of adjacent edges as formalized by the following process. From a coloring X_t, choose a pair of adjacent edges e_1, e_2 in G uniformly at random (e_1 is potentially equal to e_2) and then resample the edge-coloring on $\{e_1, e_2\}$ uniformly at random from those assignments consistent with the coloring on $G \setminus \{e_1, e_2\}$. We refer to this chain as the *heat-bath neighboring edge dynamics*.

Theorem 2. *For any tree with maximum degree Δ and n vertices, any $q \geq \Delta + 1$, then the relaxation time of the heat-bath neighboring edge dynamics for q-edge-colorings is $O_{\Delta,q}(n)$.*

We extend our optimal relaxation time bound for the Glauber dynamics with $q \geq \Delta + 2$ to a near-optimal mixing time bound on regular trees. A tree is said to be Δ-regular complete if every non-leaf vertex has degree Δ and all leaf vertices are on the same level. The root of the tree has Δ children and other non-leaf vertices have $d = \Delta - 1$ children. For the case of regular trees, Delcourt, Heinrich, and Perarnau [DHP20, Section 5] can prove $O(n^{2+o_d(1)})$ mixing time when $q \geq \Delta + 1$. We obtain near-optimal mixing time bounds for regular trees.

Theorem 3. *For all $\Delta \geq 2$, all $q \geq \Delta + 2$, for the Δ-regular complete tree with n vertices, the Glauber dynamics for q-edge-colorings has mixing time $O_{\Delta,q}(n \log^2 n)$. Moreover, for an arbitrary tree T of maximum degree Δ we obtain mixing time $O_{\Delta,q}(nD \log n)$ where D is the diameter of T.*

For regular trees we can further improve the relaxation time bound for Glauber dynamics in Theorem 3 to $q = \Delta + 1$ but at the expense of a slightly worse exponent.

Theorem 4. *For any Δ-regular complete tree with n vertices, the Glauber dynamics for $(\Delta + 1)$-edge-colorings has relaxation time $T_{rel} = \Delta n^{1+O(1/\log \Delta)}$ and mixing time $T_{mix} = (\Delta \log \Delta)^2 \cdot n^{1+O(1/\log \Delta)}$, where the big-O notation hides a universal constant.*

We prove a lower bound for the relaxation time of the Glauber dynamics in Theorem 5. This indicates that the relaxation time bound in Theorem 4 is asymptotically tight for large Δ. However, for regular trees with small degree Δ, the picture is still very mysterious: when $\Delta = 2$, there is an $\Omega(n^3)$ lower bound for the relaxation time due to [DGJ06]; when Δ is sufficiently large, Theorem 4 gives an $\Delta n^{1+o(1)}$ upper bound for the relaxation time. Based on Theorem 5, we conjecture that the Δn relaxation time holds as long as $\Delta \geq 3$.

Theorem 5. *Let $\mathbb{T}$ be a tree with n vertices and maximum degree Δ. Suppose there exists an edge e in $\mathbb{T}$ such that both endpoints of e are of degree Δ. Then, for the Glauber dynamics on q-edge-colorings with $q \in [\Delta + 1, 2\Delta]$, the relaxation time satisfies $\Omega(\Delta n / (q - \Delta)^2)$ and the mixing time satisfies $\Omega(\Delta n \log n)$.*

We provide here a very brief high-level overview of the proof of Theorem 1 which establishes an optimal relaxation time bound on the Glauber dynamics when $q \geq \Delta + 2$; a detailed overview is provided in Section 3.1. We bound the spectral gap of the Glauber dynamics by establishing a property known as approximation tensorization of variance; this is an important notion in the recent flurry of results using spectral independence, e.g., see [CLV21, BCC$^+$22].

For a graph $G = (V, E)$, consider a distribution μ on $\Omega \subset [q]^E$ where $[q] = \{1, \ldots, q\}$ for integer $q \geq 2$. If μ is a product distribution (i.e., each $e \in E$ is mutually independent of each other) then the variance of any functional $f : \Omega \to \mathbb{R}$ under the distribution μ can be upper bounded by the sum of conditional variances on each coordinate $e \in E$; this property is known as tensorization of variance [CMT15]. Our goal is to establish an approximate version of tensorization so that

2

the variance functional $\mathbf{Var}_\mu[f]$ is upper bounded by a weighted sum of the variances on each coordinate, where the weights are given by $C(e)$ for each $e \in E$. It turns out that approximate tensorization of variance with constants $C(e) \leq C$ for some constant $C \geq 1$ is equivalent to the spectral gap of the Glauber dynamics being $\geq 1/(Cn)$ (and hence the relaxation time satisfying $T_{\mathrm{rel}} \leq Cn$). We refer the reader to Definition 6 for a more formal introduction.

We establish approximate tensorization of the Δ-regular complete tree by an inductive approach. We first establish a "modified" approximate tensorization property for a Δ-regular complete tree of height $\ell = O(\Delta^2 \log^2 \Delta)$ where the constants $C(e)$ are small for the leaf edges and exponentially large (in Δ) for the internal edges. We then have an inductive approach to deduce approximate tensorization for an arbitrarily large tree, see Theorem 11 in Section 3.1 and its proof in Section 4, where the constants $C(e)$ are recursively defined, see Equation (20) in Section 4.

To establish the modified approximate tensorization property for the smaller tree of height ℓ we utilize the canonical paths approach [JS89, Sin92]. An important aspect of the proof is that we only need a good bound on the associated congestion for transitions which recolor leaf edges. The modified approximate tensorization property does not directly correspond to the relaxation time of a Markov chain, and this makes our canonical path approach somewhat different from the original version [JS89, Sin92]. In particular, our canonical path approach involves couplings, and also the overhead from the length of paths does not play a role here (see Section 3.2 for details). We believe this is the first proof utilizing the canonical paths approach in combination with approximate tensorization of variance/entropy.

Our canonical path analysis for Glauber dynamics requires $q \geq \Delta + 2$. This is mainly because our current analysis works for the $\Delta = 2$ case and the $\Delta \geq 3$ case simultaneously. A good canonical path of this type should not exist for $q = \Delta + 1$ since otherwise our proof gives $O(n)$ relaxation time for Glauber dynamics when $\Delta = 2$ and $q = 3$, which contradicts the $\Omega(n^3)$ lower bound. In practice, when $q = \Delta + 1$, we can find bad examples to fail the canonical path analysis for Glauber dynamics (see Figure 3 for details).

However, the heat-bath neighboring edge dynamics in Theorem 2 works even when $q = \Delta + 1$. When $\Delta = 2$ and $q = 3$, one can show $\widetilde{O}(n)$ mixing time for this dynamics via path coupling, which means we do not have a similar theoretical barrier as the Glauber dynamics. In practice, our canonical path analysis for the heat-bath neighboring edge dynamics heavily relies on the updates which exchange the colors of two neighboring edges if the resulting coloring is proper. Simulating this kind of update via the Glauber dynamics updates turns out to be very challenging. It still remains open whether a good collection of canonical paths exist for the Glauber dynamics when $q \geq \Delta + 1$ and $\Delta \geq 3$.

In Section 3.1 we give a detailed overview of the proof of Theorem 1 for the Glauber dynamics when $q \geq \Delta + 2$; the proofs of the main technical results are provided in Sections 4 and 5. We present an overview of the refined approach for the neighboring edge dynamics for the case $q = \Delta + 1$ in Section 3.3. The proof of Theorem 3 is also presented in Section 3.1, and the refinement from Theorem 4 is presented in Appendix B. The proof of the lower bound result in Theorem 5 is given in Appendix A.

Related works   Besides edge colorings on trees [DHP20], the mixing time and relaxation time of Glauber dynamics has been studied extensively for other Gibbs distributions on trees [MSW03, BKMP05, MSW04, LMP09, GJK10, TVVY12, RSV$^+$14, SZ17, EF23, Che24, EHSV23, CYYZ24].

Many of these results on trees [MSW03, MSW04, TVVY12, SZ17] follow the paradigm developed in [MSW03]. In these approaches, the relaxation time for a block dynamics is proved first. The block dynamics picks a vertex uniformly at random and resamples the sub-tree of some con-

stant depth beneath that vertex. Then the relaxation time for the Glauber dynamics is proved via a simple comparison argument between the block dynamics and the Glauber dynamics. We note that this approach does not suit the edge coloring problems considered here, since even the comparison between the heat-bath neighboring edge dynamics and the Glauber dynamics is non-trivial.

Recently, the Glauber dynamics on trees has been re-studied [EHSV23, CYYZ24] by using the lens of *spectral independence* [ALO20]. In fact, this framework can also be used to deduce rapid mixing of the Glauber dynamics for edge colorings on general graphs [ALG21, WZZ24] as detailed earlier. These works offer some powerful tools that can be applied to general Gibbs distributions. However, we also note that these approaches cannot apply to the edge coloring problems considered in this paper as spectral independence for arbitrary "pinnings" (see Section 2.2) does not hold in our setting as the corresponding list coloring problem may be a disconnected state space.

2 Preliminaries

For a positive integer q, let $[q] := \{1, \ldots, q\}$. Let $G = (V, E)$ be a graph. For a vertex $v \in V$, we let $N(v)$ denote the neighbors of v, and for an edge $e \in E$ we let $N(e)$ denote the edges which share an endpoint with e. Let $q \geq 2$ be an integer, denote the set of q-edge colorings on G as

$$\Omega = \Omega_G := \{\sigma \in [q]^E : \text{ for all } e \in E, f \in N(e), \sigma(e) \neq \sigma(f)\}.$$

Let μ denote the uniform distribution over $\Omega \subset [q]^E$. For any subset $S \subset E$ and any $\sigma \in \Omega$, we use μ^{σ_S} to denote the distribution of μ conditional on the configuration on S is fixed to be σ_S. Note, a random sample $X \sim \mu^{\sigma_S}$ is a full configuration on E where $X_S = \sigma_S$. Moreover, for any subset $T \subset E$, we use $\mu_T^{\sigma_S}$ to denote the marginal distribution on T projected from μ^{σ_S}.

More generally, let $\mathcal{L}_e \subseteq [q]$ be a color list for each edge $e \in E$. We can define $\mathcal{L}$-list edge-colorings for G if we further require the color of each edge e can only be chosen from its color palette $\mathcal{L}_e$. Denote the set of $\mathcal{L}$-list edge-colorings on G as

$$\Omega_{G,\mathcal{L}} := \{\sigma : \text{for all } e \in E, \sigma_e \in \mathcal{L}_e\} \cap \Omega_G.$$

We will use $\mu_{G,\mathcal{L}}$ to denote the uniform distribution over $\Omega_{G,\mathcal{L}}$. The conditional and marginal distributions can be defined accordingly.

2.1 Markov chain fundamentals

For a pair of distributions ν, π on Ω, their total variation distance is denoted as $D_{\text{TV}}(\nu, \pi) := \frac{1}{2} \sum_{\sigma \in \Omega} |\nu(\sigma) - \pi(\sigma)|$.

Consider an ergodic Markov chain $(X_t)_{t \geq 0}$ on state space Ω with transition matrix P and unique stationary distribution π. The *mixing time* T_{mix} is defined to be the minimum time, from the worst initial state X_0, to ensure that X_t is within total variation distance $\leq 1/4$ from π.

Let $1 = \lambda_1 > \lambda_2 \geq \ldots \lambda_{|\Omega|} > -1$ denote the eigenvalues of P. The *absolute spectral gap* is the quantity $1 - \lambda_*$ where $\lambda_* = \max\{|\lambda_2|, |\lambda_\Omega|\}$. Since all eigenvalues of the Glauber dynamics are non-negative [DGU14], then $\lambda_* = 1 - \lambda_2$. The inverse of the spectral gap $T_{\text{rel}} := 1/\lambda_*$ is usually known as the *relaxation time* of the Glauber dynamics. The following is a standard bound on the mixing time in terms of the relaxation time (e.g., see [LP17, Theorem 12.4]):

$$T_{\text{mix}} \leq T_{\text{rel}}(1 - \log(\min_{\sigma \in \Omega} \mu(\sigma))) \leq T_{\text{rel}}(1 + n \log q).$$

For a tree of maximum degree Δ, the Glauber dynamics on q-edge colorings is ergodic when $q \geq \Delta + 1$ [DHP20, DFFV06]. For general graphs of maximum degree Δ one needs $q \geq 2\Delta$ to guarantee ergodicity. Since the Glauber dynamics is symmetric the unique stationary distribution is the uniform distribution μ over Ω.

2.2 Approximate tensorization of variance

For a function $f : \Omega \to \mathbb{R}$, we denote the expectation and variance of f with respect to μ as follows:

$$\mu[f] = \mathbb{E}_\mu[f] := \sum_\sigma \mu(\sigma)f(\sigma) \quad \text{and} \quad \mathbf{Var}_\mu[f] = \mu[f^2] - \mu^2[f].$$

For a subset $S \subset E$, a *pinning* is an assignment $\tau : S \to [q]$. We will restrict attention to feasible pinnings which are $\tau \in [q]^S$ for some $S \subset V$ such that there exists $\sigma \in \Omega$ where $\sigma_S = \tau_S$.

For a pinning $\tau \in [q]^S$ for $S \subset E$, we use $\mu^\tau[f]$ to denote the expectation with respect to the conditional distribution μ^τ. For an edge coloring $\sigma \in \Omega$ and a subset $S \subseteq E$, denote the expectation on $E \setminus S$ with respect to the fixed assignment σ on S as

$$\mu_S[f](\sigma) := \mu^{\sigma_{E \setminus S}}[f].$$

Moreover, denote the variance with respect to the conditional distribution $\mu^{\sigma_{E \setminus S}}$ by

$$\mathbf{Var}_S[f](\sigma) := \mathbf{Var}_{\mu^{\sigma_{E \setminus S}}}[f].$$

We note that both $\mu_S[f]$ and $\mathbf{Var}_S[f]$ can also be seen as functions on $\Omega \to \mathbb{R}$.

The key property of the Gibbs distribution which implies fast mixing is known as approximate tensorization of variance.

Definition 6. We say that the distribution μ over $\Omega \subseteq [q]^E$ satisfies *approximate tensorization of variance* with constants $(C(e))_{e \in E}$ if for all $f : \Omega \to \mathbb{R}_{\geq 0}$ it holds that:

$$\mathbf{Var}[f] \leq \sum_{e \in E} C(e)\mu(\mathbf{Var}_e[f]). \tag{1}$$

If there is a universal constant γ where $C(e) \leq \gamma$ for all $e \in E$ then we say approximate tensorization of variance holds with constant γ. In words, on the RHS of eq. (1), for each edge $e \in E$, we are drawing a sample σ from μ, then fixing the configuration σ on $E \setminus \{e\}$ and looking at the variance of f with respect to the conditional distribution $\mu^{\sigma_{E \setminus e}}$. If μ is a product distribution then approximate tensorization of variance holds with constants $C(e) = 1$ for all $e \in E$, which is optimal. Variance tensorization with constants $C(e)$ bounds the relaxation time of the Glauber dynamics as $T_{\text{rel}} \leq Cn$ where $C = \max_{e \in E} C(e)$, see [CMT15, Cap23].

2.3 Relaxation time and log-Sobolev constant on regular trees

Let μ be a distribution with support $\Omega \subseteq [q]^E$. For an arbitrary non-negative function $f : \Omega \to \mathbb{R}_{\geq 0}$, we denote the entropy of f with respect to μ as follow:

$$\mathbf{Ent}_\mu[f] = \mu\left[f \log \frac{f}{\mu[f]}\right], \tag{2}$$

5

where we use the convention that $0 \log 0 = 0$. The log-Sobolev constant [DS96] for the Glauber dynamics on μ is given by

$$c_{\text{sob}} := \inf \left\{ \left. \frac{\mathbf{Ent}_\mu [f]}{\sum_{e \in E} \mu[\mathbf{Var}_e [\sqrt{f}]]} \right| f : \Omega \to \mathbb{R}_{\geq 0}, \mathbf{Ent}_\mu [f] \neq 0 \right\}. \tag{3}$$

The mixing time is related to the log-Sobolev constant c_{sob} as follows:

$$T_{\text{mix}}(\varepsilon) \leq n \cdot c_{\text{sob}}^{-1} \left(\log \log \frac{1}{\mu_{\text{min}}} + \log \frac{1}{2\varepsilon^2} \right). \tag{4}$$

We refer the interested readers to [BCP⁺22, Lemma 2.8 and Lemma 2.4] and the references therein.

As noted in [MSW03], there is a crude bound that relates the log-Sobolev constant and the spectral gap on complete Δ-regular trees. Let $\mathbb{T}_\ell$ be the complete Δ-regular tree with depth ℓ. Let μ_ℓ be the uniform distribution over all the q-edge-coloring on $\mathbb{T}_\ell$. Let $\mathbb{T}_\ell^\star$ be a d-ary tree of depth ℓ with an extra hanging root edge r. Let $\mu_\ell^{\star,c} = \mu_{\mathbb{T}_\ell^\star, \mathcal{L}}$ be the uniform distribution over all the $\mathcal{L}$-edge-list-coloring on $\mathbb{T}_\ell^\star$, where the color of the root edge r is fixed to c and $\mathcal{L}_h = [q]$ for all the other edges. We refer the reader to Definition 9 and Definition 10 for the formal definitions of $\mathbb{T}_\ell$, $\mathbb{T}_\ell^\star$ and $\mu_\ell, \mu_\ell^{\star,c}$. For $\mu \in \{\mu_\ell, \mu_\ell^{\star,c}\}$, we will use $c_{\text{sob}}(\mu)$ and $c_{\text{gap}}(\mu)$ to denote the log-Sobolev constant and the spectral gap for the continuous time Glauber dynamics on μ.

For convenience, we refer to the variant of [MSW03] in [TVVY12] which considers vertex coloring. We note that their proof also works for the edge coloring with small modifications.

Lemma 7 ([TVVY12, Lemma 26]). *For $\ell > 0$, we have the following recursions*

$$c_{\text{sob}}^{-1}(\mu_\ell^{\star,1}) \leq c_{\text{sob}}^{-1}(\mu_{\ell-1}^{\star,1}) + \alpha \cdot c_{\text{gap}}^{-1}(\mu_\ell^{\star,1}) \tag{5}$$

$$\text{and} \quad c_{\text{sob}}^{-1}(\mu_\ell) \leq c_{\text{sob}}^{-1}(\mu_{\ell-1}^{\star,1}) + \alpha \cdot c_{\text{gap}}^{-1}(\mu_\ell). \tag{6}$$

where $\alpha = \frac{\log(k-2)}{1-2/(k-1)}$ with $k = q^\Delta$ and we use $c_{\text{sob}}^{-1}(\mu_0) = c_{\text{sob}}^{-1}(\mu_0^{\star,1}) = 1$ as conventions.

We note that Lemma 7 is a slight generalization from the version in [TVVY12]. In particular, eq. (5) corresponds to the statement in [TVVY12]; here every tree has a pinned parent of the root. We also consider the case where the root does not have a parent, see eq. (6); this inequality can be proved by the same proof as in [TVVY12].

Corollary 8. *For $\ell > 0$, it holds that*

$$c_{\text{sob}}^{-1}(\mu_\ell) \leq \alpha \ell \cdot \max \left\{ c_{\text{gap}}^{-1}(\mu_\ell), \max_{1 \leq j < \ell} c_{\text{gap}}^{-1}(\mu_j^{\star,1}) \right\}. \tag{7}$$

Corollary 8 corrects a slight error appearing in [TVVY12, Proof of Theorem 6]. The analog of Equation (7) in [TVVY12] is missing the factor ℓ on the right-hand side; however, this appears to be due to a mismatch of continuous-time and discrete-time chain statements in [TVVY12, Lemmas 26 and 27]. Consequently, [TVVY12] use $O(n)$ relaxation time of the complete regular tree to deduce $O(n \log n)$ mixing time, but we believe there is an extra factor of $O(\log n)$ that appears in this mixing time bound.

3 Relaxation time analysis

In this section we will focus on establishing a tight upper bound on the relaxation time for Δ-regular complete trees when $q \geq \Delta + 2$. We will then use this to establish Theorem 1 which yields $O(n)$ relaxation time for the Glauber dynamics on any tree of maximum degree Δ when $q \geq \Delta + 2$. Then using a refinement of the approach presented in this section we will prove Theorem 2 (in Section 3.3) which yields $O(n)$ relaxation time for an edge dynamics when $q = \Delta + 1$.

We prove the upper bound on the relaxation time by establishing approximate tensorization of variance (see Definition 6) with a constant $C = C(\Delta)$; recall this implies $T_{\text{rel}} = O(n)$ for the Glauber dynamics.

We will consider the more general setting of list edge colorings. Given a tree $\mathbb{T} = (V, E)$ and a color list $\mathcal{L}_e \subseteq [q]$ for each edge $e \in E$, let $\mu_{\mathbb{T},\mathcal{L}}$ denote the uniform distribution of all proper list edge colorings, where a list edge coloring assigns each edge e a color from $\mathcal{L}_e$ such that adjacent edges receive different colors. We often use μ to denote $\mu_{\mathbb{T},\mathcal{L}}$ if $\mathbb{T}$ and $\mathcal{L}$ are clear from the context.

There are two Δ-regular trees that will play a key role in our analysis. Let $\mathbb{T}_k$ denote the complete Δ-regular tree with k edge levels (we call k the depth of the tree), where the root has Δ children and other non-leaf vertices have $d = \Delta - 1$ children. All edges incident to the root are in level 1. For $1 \leq i \leq k$, we use $L_i(\mathbb{T}_k)$ to denote all edges in the i-th level, which is the set of edges with distance $i - 1$ to $L_1(\mathbb{T}_k)$ in the line graph. Let $\mu_k = \mu_{\mathbb{T}_k,\mathcal{L}}$ to denote the uniform distribution of all proper list edge colorings where $\mathcal{L}_e = [q]$ for all $e \in E$.

Let $\mathbb{T}_k^\star$ denote the tree formed by taking the complete d-ary tree of depth k (all non-leaf vertices, including the root, have $d = \Delta - 1$ children) and adding an additional edge r adjacent to the root vertex. In this tree, every non-leaf vertex has degree Δ, and there is one additional hanging root edge $r = \{u, v\}$ incident to the root where u is the root and v is an additional (leaf) vertex. As before, we can define all sets $L_i(\mathbb{T}_k^\star)$ for $1 \leq i \leq k$ and let $L_0(\mathbb{T}_k^\star) = \{r\}$. Let $\mu_k^\star = \mu_{\mathbb{T}^\star,\mathcal{L}}$ denote the uniform list edge-coloring distribution for $\mathbb{T}_\ell^\star$, where $\mathcal{L}_r = [q - d]$ and $\mathcal{L}_h = [q]$ for other edges $h \neq r$. Moreover, we denote $\mu_k^{\star,c}$ when the color of the hanging root edge r is fixed to color $c \in [q]$. Since this is symmetric over the colors, we simply consider $\mu_k^{\star,1}$.

The following definition summarizes the above description of the two kinds of Δ-regular trees which will be relevant in this paper.

Definition 9 ($\mathbb{T}_k, \mathbb{T}_k^\star$). $\mathbb{T}_k$ is the tree that has Δ children for the root u and then branches with factor $d = \Delta - 1$ until level k; whereas $\mathbb{T}_k^\star$ has an extra hanging root edge $r = (u, v)$ where v is an additional leaf vertex and then branches with factor $d = \Delta - 1$ for all non-leaf nodes until level k.

Then, on $\mathbb{T}_k$ and $\mathbb{T}_k^\star$, there are three specific distributions we consider.

Definition 10 ($\mu_k, \mu_k^\star, \mu_k^{\star,1}$). Suppose $q \geq \Delta + 1$ is fixed, we have:

- $\mu_k = \mu_{\mathbb{T}_k,\mathcal{L}_1}$, where $\mathcal{L}_1(e) = [q]$ for all e;

- $\mu_k^\star = \mu_{\mathbb{T}_k^\star,\mathcal{L}_2}$, where $\mathcal{L}_2(r) = [q - d]$ and $\mathcal{L}_2(e) = [q]$ for $e \neq r$;

- $\mu_k^{\star,1} = \mu_{\mathbb{T}_k^\star,\mathcal{L}_3}$, where $\mathcal{L}_3(r) = \{1\}$ and $\mathcal{L}_3(e) = [q]$ for $e \neq r$.

3.1 Proof overview for Theorems 1 and 3: Glauber dynamics for $q \geq \Delta + 2$

Consider a tree $T = (V, E)$ of maximum degree Δ. For integer $q \geq \Delta + 1$, let μ denote a distribution over a subset of $[q]^E$. (We will use μ to be either μ_ℓ, $\mu_\ell^\star$, or $\mu_\ell^{\star,1}$) Recall, *approximate tensorization of*

variance for μ which holds with constants $C(e), e \in E$ if for all $f : \Omega \to \mathbb{R}_{\geq 0}$:

$$\mathbf{Var}_\mu [f] \leq \sum_{e \in E} C(e) \mu(\mathbf{Var}_e [f]). \tag{1}$$

We will also consider the following modified notion. Consider the tree $\mathbb{T}_k^\star$ and let $r = (u, v)$ denote the hanging root edge. We say that *approximate* **root**-*tensorization of variance* for $\mu_k^\star$ holds with constants $C(e), e \in E$ if for all $f : \Omega \to \mathbb{R}_{\geq 0}$:

$$\mathbf{Var}_{\mu_k^\star} \left[\mu_{\mathbb{T}_k^\star \setminus \{r\}}[f] \right] \leq \sum_{e \in E} C(e) \mu_k^\star(\mathbf{Var}_e [f]). \tag{8}$$

Note, $\mathbf{Var}_{\mu_k^\star}[\mu_{\mathbb{T}_k^\star \setminus \{r\}}[f]] = \mathbf{Var}_{\sigma \sim \mu_k^\star}[\mathbb{E}_{\tau \sim \mu_k^{\star,\sigma(r)}} [f(\tau)]]$. In words, in the LHS of eq. (8) we look at the variance of the following quantity X: choose a sample $\sigma \sim \mu_k^\star$, fix the color $\sigma(r)$ of the hanging root edge r, and then let X be the expectation of $f(\tau)$ where $\tau \sim \mu_k^{\star,\sigma(r)}$.

Theorem 11. *Let $\Delta \geq 2$, $q \geq \Delta + 1$ be two integers. Suppose there exists an integer $\ell \geq 1$ where the following holds:*

1. *There exist constants $\alpha = (\alpha_0, \ldots, \alpha_\ell)$ and approximate **root**-tensorization of variance for $\mu_\ell^\star$ holds with constants $C(e) = \alpha_i$ where $e \in L_i(\mathbb{T}_\ell^\star)$;*

2. *There exists a constant γ such that for $1 \leq j \leq \ell$, both μ_j and $\mu_j^{\star,1}$ satisfy approximate tensorization of variance with constant γ.*

Then, for any $k \geq 1$, μ_k and $\mu_k^{\star,1}$ satisfy approximate tensorization of variance with constants

$$C'(e) \leq \gamma \cdot \left(\max_{0 \leq j \leq \ell} \alpha_j \cdot \sum_{i=0}^{\lfloor k/\ell \rfloor} \alpha_\ell^i + \max \{1, \alpha_0\} \right).$$

Note that ℓ can depend on Δ and q, and α and γ can depend on ℓ (and hence Δ and q). The proof of Theorem 11 is contained in Section 4.

Corollary 12. *Under the same conditions as Theorem 11, if $\alpha_\ell < 1$ then the relaxation time of the Glauber dynamics is $O_{\alpha,\gamma}(n)$ on $\mathbb{T}_k$, the Δ-regular tree of depth k.*

We now give an overview on proving Theorem 1 which gives an optimal relaxation time bound for the Glauber dynamics on trees of maximum degree Δ when $q \geq \Delta + 2$. First, we use Corollary 12 to prove an $O_{\Delta,q}(n)$ relaxation time bound for the Glauber dynamics on $\mathbb{T}_k$. The key step is to establish the conditions in Theorem 11.

Lemma 13. *Let $\Delta \geq 2$, $q = \Delta + 2$ be two integers. There exists $\ell = O(\Delta^2 \log^2 \Delta)$ such that the constants $\alpha = (\alpha_0, \ldots, \alpha_\ell)$ and γ in Theorem 11 exist with $\alpha_\ell = \frac{1}{2}$, $\alpha_j = q^{\Delta^{O(\ell)}}$ for all $j < \ell$ and $\gamma = q^{\Delta^{O(\ell)}}$.*

Combining Corollary 12 with Lemma 13, one can immediately obtain the relaxation time for Δ-regular trees. To prove Theorem 1, we need to explain how the relaxation time of Δ-regular trees implies the relaxation time of arbitrary trees of maximum degree Δ. As noted by the previous work of [DHP20], there is a comparison argument of the relaxation time between the Glauber dynamics on the Δ-regular tree and arbitrary trees. This comparison argument for the relaxation time can be easily extend to the log-Sobolev constant c_{sob} for continuous time Glauber dynamics (see Section 2.3 for a definition).

Lemma 14 ([DHP20, Proposition 9]). *Let $\mathbb{T} = (V, E)$ be a tree with root r and maximum degree Δ. Let $\mathbb{T}_H = (V_H, H \subseteq E)$ be a connected sub-tree of $\mathbb{T}$ containing r. Let $q \geq \Delta + 1$, and μ, $\nu = \mu_H$ be the distribution of uniform q-edge-coloring on $\mathbb{T}$ and $\mathbb{T}_H$, respectively. Then,*

- *μ satisfies approximate tensorization of variance with constant C implies ν satisfies approximate tensorization of variance with constant Cq;*

- *the log-Sobolev constant on ν is bounded by $c_{\mathrm{sob}}^{-1}(\nu) \leq q \cdot c_{\mathrm{sob}}^{-1}(\mu)$.*

The proof of Lemma 14 could be done by a brute-force comparison between each term of (1) for μ and ν. Since the notation used here is quite different from the one in [DHP20], for completeness, we include a proof of Lemma 14 in Appendix D.

We can now prove Theorems 1 and 3.

Proof of Theorem 1. Given a tree $\mathbb{T}$ with maximum degree Δ and $q \geq \Delta + 2$. Let ν be the distribution of uniform q-edge-coloring on $\mathbb{T}$. We consider the distribution of uniform q-edge-coloring μ on the $(q-2)$-regular tree $\mathbb{T}_k$ for sufficiently large depth k and degree $\Delta' = q - 2$ such that $\mathbb{T}$ is a subtree of $\mathbb{T}_k$. Since $q \geq \Delta + 2$, we have $\Delta' \geq \Delta$ and such $\mathbb{T}_k$ must exist.

By Corollary 12 and Lemma 13, we know that when $q = \Delta' + 2$, the relaxation time of the Glauber dynamics on μ is $O_{\Delta', q}(n) = O_q(n)$. This gives $O_q(1)$ approximate tensorization of variance of μ. Then, after applying Lemma 14, we know that ν also has $O_q(1)$ approximate tensorization of variance. This proves the $O_{\Delta, q}(n)$ relaxation time for the Glauber dynamics on ν. □

Proof of Theorem 3. Let $\mathbb{T}$ be a tree with maximum degree Δ and $q \geq \Delta + 2$ and let ν be the distribution of uniform q-edge-colorings on $\mathbb{T}$. We consider the distribution of uniform q-edge-coloring μ on the $(q-2)$-regular tree $\mathbb{T}_k$ with maximum degree $\Delta' = q - 2$ and sufficiently large depth k. In particular, we choose the depth k according to the following rules:

- when $\mathbb{T}$ is a complete Δ-regular tree of depth ℓ (this means $\ell = O(\log_\Delta n)$), we let $k = \ell$;

- otherwise we let $k = D$ to be the diameter of $\mathbb{T}$.

Since $q \geq \Delta + 2$, we have $\Delta' \geq \Delta$. This means in both cases, $\mathbb{T}$ is a subtree of $\mathbb{T}_k$.

By Theorem 11 and Lemma 13, when $q = \Delta' + 2$, we know that for any $j \geq 1$, both μ_j and $\mu_j^{\star,1}$ satisfy approximate tensorization of variance with constants $O_{\Delta', q}(1) = O_q(1)$. This implies the relaxation time (or, the inverse of spectral gap) of the continuous time Glauber dynamics on μ_j and $\mu_j^{\star,1}$ are also $O_q(1)$. That is, for any $j \geq 1$, we have

$$c_{\mathrm{gap}}^{-1}(\mu_j) = O_q(1) \quad \text{and} \quad c_{\mathrm{gap}}^{-1}(\mu_j^{\star,1}) = O_q(1).$$

Combining above equations with Corollary 8, for any $\ell \geq 1$, it holds that $c_{\mathrm{sob}}^{-1}(\mu) = c_{\mathrm{sob}}^{-1}(\mu_k) = k \cdot O_q(1)$. By Lemma 14, the log-Soblev constant for ν on $\mathbb{T}$ is bounded by $c_{\mathrm{sob}}^{-1}(\nu) \leq q c_{\mathrm{sob}}^{-1}(\mu) = k \cdot O_q(1)$. Finally, according to the standard relation between log-Sobolev constant and the mixing time described in (4), it holds that the mixing time for the Glauber dynamics on ν is bounded by

$$T_{\mathrm{mix}} = O_q(1) \cdot k \cdot n \log n.$$

Note that by our choice of k, we have $k = O(\log_\Delta n)$ when $\mathbb{T}$ is a regular complete tree; and $k = D$ be the diameter of $\mathbb{T}$ when $\mathbb{T}$ is an arbitrary tree. This finishes the proof. □

3.2 Relaxation time and root-tensorization for complete regular tree

In this section, we will outline the proof of Lemma 13. In Lemma 13, we use a brute-force bound for all α_j with $0 \leq j < \ell$ and γ. The most difficult part is to guarantee $\alpha_\ell = \frac{1}{2} < 1$ so that we can apply Corollary 12 to obtain Theorem 1. In this subsection, we explain how to establish the *first condition* in Theorem 11, namely, the existance of constants $\alpha = (\alpha_0, \ldots, \alpha_\ell)$ with $\alpha_\ell < 1$.

In Lemma 13, we assume $q = \Delta + 2$. We first introduce some simplified notation that will be used in this subsection. Since we focus on constants α, we only care about the tree $\mathbb{T}_\ell^\star$ (see Definition 9), and the list coloring distribution $\mu_\ell^\star$ in Theorem 11. Recall that $\mu_\ell^\star$ is the uniform distribution of list colorings in $\mathbb{T}_\ell^\star$, where the root hanging edge has the color list $[q - d] = [3]$ and all other edges have color list $[q]$. For simplicity of notation, we drop the index ℓ and denote $\mathbb{T}^\star = \mathbb{T}_\ell^\star$. Also, let $\widetilde{\mathbb{T}} := \mathbb{T}^\star - r$. We use μ to denote the distribution $\mu_\ell^\star$.

Our goal is to establish approximate root-tensorization of variance as defined in (8). Recall this states that for any function $f : \Omega(\mu) \to \mathbb{R}$, the following holds:

$$\mathbf{Var}_\mu \left[\mu_{\widetilde{\mathbb{T}}}[f] \right] \leq \sum_{i=0}^{\ell} \alpha_i \sum_{h \in L_i} \mu[\mathbf{Var}_h [f]], \tag{9}$$

where we use L_i to denote $L_i(\mathbb{T}^\star)$ to simplify notation.

Let us expand the LHS of (9). For a color $a \in [q]$, let μ^{ra} denote the conditional distribution $\mu^{r \leftarrow a}$. From the definition of variance we have the following identity:

$$\mathbf{Var}_\mu \left[\mu_{\widetilde{\mathbb{T}}}[f] \right] = \frac{1}{2} \sum_{a,b \in [q-d]:a \neq b} \mu_r(a)\mu_r(b) \left(\mu^{ra}[f] - \mu^{rb}[f] \right)^2. \tag{10}$$

Our technique for bounding $\mathbf{Var}_\mu \left[\mu_{\widetilde{\mathbb{T}}}[f] \right]$, using the decomposition in (10), is inspired by two classical tools: coupling and canonical paths. However, we use them in a non-standard manner. For two colors $a, b \in [q - d]$ with $a \neq b$, let $\mathcal{C}_{ab}$ denote a coupling of the distributions μ^{ra} and μ^{rb}. Formally, for any random sample $(\sigma, \tau) \sim \mathcal{C}_{ab}$, it holds that $\sigma \sim \mu^{ra}$ and $\tau \sim \mu^{rb}$. Let $\Omega(\mathcal{C}_{ab}) \subseteq \Omega(\mu^{ra}) \times \Omega(\mu^{rb})$ denote the support of the coupling $\mathcal{C}_{ab}$. Let $\mathcal{C} = \{\mathcal{C}_{ab} \mid a, b \in [q - d], a \neq b\}$ denote the set of couplings and let $\Omega(\mathcal{C}) = \bigcup_{\mathcal{C}_{ab} \in \mathcal{C}} \Omega(\mathcal{C}_{ab})$ denote the union of their supports. For two edge colorings γ and γ', we will use $\gamma \oplus \gamma' := \{e : \gamma_e \neq \gamma'_e\}$ to denote the set of edges on which γ and γ' are different.

Given a set of couplings, we will define the following collection of canonical paths. Let $\Lambda(\mu)$ denote the ordered pairs $(\gamma, \gamma') \in \Omega(\mu)$ where $|\gamma \oplus \gamma'| = 1$, i.e., $\Lambda(\mu)$ are the pairs of states with positive probability in the transition matrix for the Glauber dynamics.

Definition 15 (Canonical paths w.r.t. a coupling set). Given a coupling set $\mathcal{C}$, for any pair $(\sigma, \tau) \in \Omega(\mathcal{C})$, a canonical path $\gamma^{\sigma,\tau}$ is a simple path from σ to τ in the graph $(\Omega(\mu), \Lambda(\mu))$. More specifically, the path $\gamma^{\sigma,\tau} = (\gamma_0, \gamma_1, \ldots, \gamma_m)$ with the following properties: (1) $\sigma = \gamma_0$, $\tau = \gamma_m$ and all γ_i's are distinct; (2) for any $1 \leq i \leq m$, γ_{i-1} and γ_i differ only at the color on one edge.

Let $\Gamma = \{\gamma^{\sigma,\tau} \mid (\sigma, \tau) \in \Omega(\mathcal{C})\}$ denote the set of canonical paths w.r.t. the coupling set $\mathcal{C}$.

In the following we use $Q_\mu^{\text{Glauber}}(\gamma, \gamma')$ to denote the transition rate from γ to γ' in the (continuous-time) heat-bath Glauber dynamics for μ. Formally,

$$Q_\mu^{\text{Glauber}}(\gamma, \gamma') = \mu^{\gamma_{\mathbb{T}\backslash e}}(\gamma') = \mu_e^{\gamma_{\mathbb{T}\backslash e}}(\gamma'_e).$$

This corresponds to the transition probability from γ to γ' for the heat-bath (discrete-time) Glauber dynamics conditional on the edge $\gamma \oplus \gamma'$ is recolored.

We define the following quantity as the expected congestion of the canonical paths Γ.

Definition 16 (Congestion of canonical paths w.r.t. a coupling set)**.** Given a coupling set $\mathcal{C}$ and a set of canonical paths Γ, define the expected congestion at level $0 \leq t \leq \ell$ by

$$\xi_t = \max_{\substack{a,b \in [q-d]: \\ a \neq b}} \sum_{\substack{s=(\gamma,\gamma') \in \Lambda(\mu): \\ (\gamma \oplus \gamma') \in L_t}} \frac{\left(\mathbf{Pr}_{(\sigma,\tau) \sim \mathcal{C}_{ab}} \left[s \in \gamma^{\sigma,\tau} \right] \right)^2}{\mu(\gamma) Q_\mu^{\text{Glauber}}(\gamma, \gamma')}. \tag{11}$$

The congestion in (11) sums over all pairs $(\gamma, \gamma') \in \Omega(\mu) \times \Omega(\mu)$ that differ only at one edge in level L_t. The numerator is the square of the probability that a random canonical path $\gamma^{\sigma,\tau}$, where $(\sigma, \tau) \sim \mathcal{C}_{ab}$, uses (γ, γ'). The denominator is the capacity of (γ, γ') in the (continuous-time) Glauber dynamics.

Remark 17 (Comparison to standard canonical paths approach)**.** The canonical paths technique is a classical tool for upper bounding the mixing time of Markov chains, see [JS89, DS91, Sin92]. In the standard approach, we define a canonical path $\gamma_{\text{classical}}^{\sigma,\tau}$ between *every pair* of colorings $\sigma, \tau \in \Omega(\mu)$. The goal is to define canonical paths which minimize the congestion defined to be the maximum "load" through any transition. The congestion then yields a bound on the spectral gap of the Glauber dynamics (equivalently, it establishes approximate tensorization of variance).

The standard canonical paths technique does not involve any coupling. Here, our purpose is to establish approximate **root**-tensorization of variance. Our canonical path is constructed with respect to a set of couplings $\mathcal{C}$: for the coupling $\mathcal{C}_{ab}$, we are going from a coloring σ where the root edge r has a color a to a coloring τ where the root edge r has color b, such that $(\sigma, \tau) \sim \mathcal{C}_{ab}$.

The congestion [Sin92] in the classical canonical paths approach is defined as follows:

$$\rho := \max_{(\gamma,\gamma')} \frac{1}{\mu(\gamma) Q_\mu^{\text{Glauber}}(\gamma, \gamma')} \sum_{\substack{(\sigma,\tau) \in \Lambda(\mu): \\ (\gamma,\gamma') \in \gamma_{\text{classical}}^{\sigma,\tau}}} \mu(\sigma) \mu(\tau) |\gamma_{\text{classical}}^{\sigma,\tau}|,$$

where $|\gamma_{\text{classical}}^{\sigma,\tau}|$ is the length of the path $\gamma_{\text{classical}}^{\sigma,\tau}$. Comparing ρ with our ξ_t, there are some important differences. (1) Our definition of congestion involves the probability of a pair in the coupling $\mathcal{C}$. (2) We measure the total congestion in a level rather than the congestion in a single transition. (3) Our congestion does not contain the length of the path, this is because when we use ξ_t to establish (9), we apply the Cauchy-Schwarz inequality in a different manner than in [Sin92], see the proof of Lemma 18 for more details.

The most important difference is in our application of the technique. To obtain a tight bound on the mixing time we need $\xi_\ell < 1/(\ell + 1)$ (which will imply $\alpha_\ell < 1$) and we only need a trivial bound for levels $t < \ell$, see Lemma 13.

Finally, if we construct a set of canonical paths w.r.t. a coupling set $\mathcal{C}$ which has congestion $\boldsymbol{\xi} = (\xi_0, \xi_1, \ldots, \xi_\ell)$ then we obtain approximate root-tensorization of variance with constant $\boldsymbol{\alpha} = (\alpha_0, \ldots, \alpha_\ell)$ where $\alpha_i = (\ell + 1)\xi_i$.

Lemma 18. *If there exist a set of couplings $\mathcal{C}$ and a set Γ of canonical paths such that the congestion with respect to the coupling $\mathcal{C}$ is ξ_t for $0 \leq t \leq \ell$, then (9) holds with $\boldsymbol{\alpha} = (\ell + 1)\boldsymbol{\xi}$.*

Proof. We note that by definition

$$\mathbf{Var}_\mu \left[\mu_{\widetilde{\mathbb{T}}}[f] \right] = \frac{1}{2} \sum_{a,b \in [q-d]: a \neq b} \mu_r(a) \mu_r(b) \left(\mu^{ra}(f) - \mu^{rb}(f) \right)^2$$

$$\text{(by coupling)} \quad = \frac{1}{2} \sum_{a,b \in [q-d]: a \neq b} \mu_r(a) \mu_r(b) \left(\mathbb{E}_{(\sigma,\tau) \sim \mathcal{C}_{ab}} [f(\sigma) - f(\tau)] \right)^2$$

$$\text{(by telescoping sum)} \quad = \frac{1}{2} \sum_{a,b \in [q-d]: a \neq b} \mu_r(a) \mu_r(b) \left(\mathbb{E}_{(\sigma,\tau) \sim \mathcal{C}_{ab}} \left[\sum_{j=1}^{m(\sigma,\tau)} f(\gamma_j^{\sigma\tau}) - f(\gamma_{j-1}^{\sigma\tau}) \right] \right)^2.$$

Fix a and b. Denote $\mathbb{E}_{(\sigma,\tau) \sim \mathcal{C}_{ab}} \left[\sum_{j=1}^{m(\sigma,\tau)} f(\gamma_j^{\sigma\tau}) - f(\gamma_{j-1}^{\sigma\tau}) \right]$ by E_{ab}. Then E_{ab} can be rewritten as

$$E_{ab} = \sum_{s=(\gamma,\gamma') \in \Omega(\mu)} (f(\gamma) - f(\gamma')) \mathbf{Pr}_{(\sigma,\tau) \sim \mathcal{C}_{ab}} [s \in \gamma^{\sigma,\tau}], \quad (12)$$

where the sum is over all *ordered* pairs (γ, γ') that differ at one edge and the equation uses the fact that every canonical path uses (γ, γ') at most once. We can partition all (γ, γ')'s according to the level of the edge $e = \gamma \oplus \gamma'$. By using Cauchy-Schwarz,

$$E_{ab}^2 \leq (\ell+1) \sum_{i=0}^{\ell} \left(\sum_{\substack{s=(\gamma,\gamma') \in \Lambda(\mu): \\ \gamma \oplus \gamma' \in L_i}} (f(\gamma) - f(\gamma')) \mathbf{Pr}_{(\sigma,\tau) \sim \mathcal{C}_{ab}} [s \in \gamma^{\sigma,\tau}] \right)^2 := (\ell+1) \sum_{i=0}^{\ell} A_i^2, \quad (13)$$

where we use A_i^2 to denote the term for each level i. Let $Q(\gamma, \gamma')$ denote $\mu(\gamma) Q_\mu^{\text{Glauber}}(\gamma, \gamma')$. By applying Cauchy-Schwarz on A_i^2 we obtain the following bound:

$$A_i^2 = \left(\sum_{\substack{s=(\gamma,\gamma') \in \Lambda(\mu): \\ \gamma \oplus \gamma' \in L_i}} (f(\gamma) - f(\gamma')) \sqrt{Q(\gamma, \gamma')} \frac{\mathbf{Pr}_{(\sigma,\tau) \sim \mathcal{C}_{ab}} [s \in \gamma^{\sigma,\tau}]}{\sqrt{Q(\gamma, \gamma')}} \right)^2$$

$$\leq \sum_{\substack{s=(\gamma,\gamma') \in \Lambda(\mu): \\ \gamma \oplus \gamma' \in L_i}} (f(\gamma) - f(\gamma'))^2 Q(\gamma, \gamma') \sum_{\substack{s=(\gamma,\gamma') \in \Lambda(\mu) \\ \gamma \oplus \gamma' \in L_i}} \frac{\left(\mathbf{Pr}_{(\sigma,\tau) \sim \mathcal{C}_{ab}} [s \in \gamma^{\sigma,\tau}] \right)^2}{Q(\gamma, \gamma')}$$

$$\leq 2\xi_i \sum_{e \in L_i} \mu[\mathbf{Var}_e [f]].$$

Combining everything together, we have

$$\mathbf{Var}_\mu \left[\mu_{\widetilde{\mathbb{T}}}[f] \right] \leq \sum_{a,b \in [q-d]: a \neq b} \mu_r(a) \mu_r(b) (\ell+1) \sum_{i=0}^{\ell} \xi_i \sum_{e \in L_i} \mu[\mathbf{Var}_e [f]] \leq (\ell+1) \sum_{i=0}^{\ell} \xi_i \sum_{e \in L_i} \mu[\mathbf{Var}_e [f]].$$

Hence, approximate root-tensorization of variance in (9) holds with $\alpha = (\ell+1)\xi$. $\square$

Using the technique developed above, we prove Lemma 13 via the following result.

Lemma 19. *Let $\Delta \geq 2$, $q = \Delta + 2$ be two integers. There exists $\ell = O(\Delta^2 \log^2 \Delta)$ such that the following holds for the uniform distribution $\mu = \mu_\ell^\star$ of the list colorings on $\mathbb{T}^\star = \mathbb{T}_\ell^\star$. There exists a set of couplings $\mathcal{C}$ and canonical paths Γ w.r.t. $\mathcal{C}$ such that the congestion ξ satisfies $\xi_\ell = \frac{1}{2(\ell+1)}$ and $\xi_j = q^{\Delta^{O(\ell)}}$ for $j < \ell$.*

The high-level idea of the canonical path construction and the analysis of the associated congestion used in the proof of Lemma 19 is the following (see Section 5 for more details). We first define, for a pair of colors $a, b \in [q]$, the coupling $\mathcal{C}_{ab}$ between a coloring σ with the root edge having color $\sigma(r) = a$ and a coloring τ with $\tau(r) = b$. To define this coupling, given σ we consider the (a, b)-colored alternating path $\mathcal{E}^*$ from the root edge r, and we define τ as the coloring obtained by flipping $\mathcal{E}^*$, which means we interchange the colors a and b on $\mathcal{E}^*$; this results in a valid coloring since $\mathcal{E}^*$ is a maximal 2-colored component.

Now consider a pair of colorings σ and τ which only differ by a (a, b)-colored alternating path $\mathcal{E}^*$ from the root edge r. Let $\mathcal{E}^* = (e_0, e_1, \ldots, e_s)$ where $e_0 = r$ and the edges are in order from the root towards the leaves. In the first stage we aim to recolor the odd edges from $\sigma(e_i) = b$ to a color c_i different than a and b. To do that we recolor a path $\mathcal{E}_i$, which is defined starting from a sibling of e_i towards the leaves. By recoloring $\mathcal{E}_i$ with particular colors we will free a color $c_i \neq a, b$ for edge e_i. However, we recolor the edges of $\mathcal{E}_i$ and the odd edges of $\mathcal{E}^*$ in the following manner. Let $\mathcal{E}$ denote the set of all edges recolored in this stage. We order the edges of $\mathcal{E}$ so that edges on lower levels (closer to the leaves) are recolored before edges on higher levels of the tree, and within a level the edges of the alternating path $\mathcal{E}^*$ are last. We then recolor the edges of $\mathcal{E}$ in this order.

In the second stage we recolor the even edges of $\mathcal{E}^*$ from color a to color b, which is available due to the first stage. Finally in the last stage we recolor the edges of $\mathcal{E}$ to their color in τ, and we do these recolorings in reverse of the ordering used in the first stage. The key to the congestion analysis is that we only need a non-trivial bound on the congestion for transitions which recolor leaf edges. This is why in the first stage we first recolor all leaf edges before proceeding up the tree, and in the last stage we recolor the leaf edges last. When calculating the congestion on leaf edges, a specific transition on a leaf edge can still be used by exponentially large number of canonical paths. However, such large number of possibilities will be offset by the exponentially small probability of the path $\mathcal{E}_i$ reaching the leaves.

3.3 Proof overview for Theorem 2: Heat-bath edge dynamics for $q = \Delta + 1$

In this section, we will prove Theorem 2. Recall, approximate tensorization of variance implies optimal relaxation time for the Glauber dynamics. Approximate tensorization can be generalized to a block factorization, which implies fast mixing for the more general block dynamics [CP21].

Definition 20. We say that the distribution μ over $\Omega \subseteq [q]^E$ satisfies *block factorization of variance* with constants $(C(B))_{B \subseteq E}$ if for all $f : \Omega \to \mathbb{R}_{\geq 0}$ it holds that:

$$\mathbf{Var}[f] \leq \sum_{B \subseteq E} C(B)\mu(\mathbf{Var}_B[f]). \tag{14}$$

For convenience we denote the support of C as $\mathcal{B} := \{B \subseteq E \mid C(B) > 0\}$.

In particular, for $\mathcal{B} = \{\{e\} \mid e \in E\}$, which is the set of all the singleton edges, then block factorization becomes approximate tensorization.

Consider the the heat-bath block dynamics P_B defined by a collection of constants $(C(B))_{B \subseteq E}$ from Definition 20. It updates a configuration X_t to X_{t+1} as follows:

1. pick a block $B \subseteq E$ with probability proportional to $C(B)$;

2. sample $X_{t+1} \sim \mu(\cdot \mid X_t(E \setminus B))$.

Then (14) immediately implies an upper bound on the relaxation time of P_B as $T_{\text{rel}}(P_B) \leq \sum_{B \subseteq E} C(B)$.

We recall the definition of the heat-bath edge dynamics. Let $\mathbb{T} = (V, E)$ be a tree with maximum degree Δ and let

$$\mathcal{P} = \mathcal{P}(\mathbb{T}) := \left\{ \{e\} \mid e \in E \right\} \cup \left\{ \{e, e'\} \mid e, e' \text{ are adjacent edges in } \mathbb{T} \right\}. \tag{15}$$

Also, let $q \geq \Delta + 1$ and let μ be the uniform distribution of q-edge-coloring on $\mathbb{T}$. Then, the heat-bath neighboring edge dynamics is exactly the heat-bath block dynamics defined by constants

$$C(B) = \mathbf{1}[B \in \mathcal{P}], \text{ for } B \in \mathcal{P}.$$

In order to prove $O_{\Delta,q}(n)$ relaxation time of the heat-bath edge dynamics, it is sufficient to prove the following tensorization bound of variance, for every $f : \Omega(\mu) \to \mathbb{R}$, there is a constant $C = C(\Delta, q)$ such that:

$$\mathbf{Var}_\mu[f] \leq C \sum_{S \in \mathcal{P}} \mu[\mathbf{Var}_S[f]]. \tag{16}$$

For convenience, if (16) holds, then we say μ satisfies *approximate edge-tensorization* with constant C. We will prove (16) with a similar high-level plan as in the proof of Theorem 1. All we need is a slightly generalized version of Theorem 11 which enables us to establish block factorization of variance with support $\mathcal{P}$ on regular trees. First, we generalize (8) by allowing general blocks rather than singletons. Let $\mathcal{B}$ be a collection of edge sets defined as the following:

$$\mathcal{B} := \left\{ \{e\} \mid e \in \mathbb{T}_\ell^\star \right\} \cup \left\{ \{e, r\} \mid e \in L_1(\mathbb{T}_\ell^\star) \right\}. \tag{17}$$

In words, $\mathcal{B}$ contains all the singleton edges in $\mathbb{T}_\ell^\star$ and the adjacent edge pairs which contain the root hanging edge r. We say that *approximate **root**-factorization of variance* for $\mu_k^\star$ holds with constants $C(B)$ if for all $f : \Omega \to \mathbb{R}_{\geq 0}$:

$$\mathbf{Var}_{\mu_k^\star}\left[\mu_{\mathbb{T}_k^\star \setminus \{r\}}[f] \right] \leq \sum_{B \in \mathcal{B}} C(B) \mu_k^\star(\mathbf{Var}_B[f]). \tag{18}$$

Theorem 21. *Let $\Delta \geq 2$, $q \geq \Delta + 1$ be two integers. Suppose there exists an integer $\ell \geq 2$ where the following holds:*

- *There exist constants $\alpha = (\alpha_0, \ldots, \alpha_\ell)$ and β and approximate **root**-factorization of variance for $\mu_k^\star$ holds with constants $C(\{e\}) = \alpha_i$ where $e \in L_i(\mathbb{T}_\ell^\star)$ and $C(B) = \beta$ where $|B| = 2$;*

- *There exists a constant γ such that for $1 \leq j \leq \ell$, both μ_j and $\mu_j^{\star,1}$ satisfy approximate tensorization of variance with constant γ.*

Then, for any $k \geq 1$, μ_k satisfies block factorizaiton of variance with constants

$$C'(B) \leq \begin{cases} \beta\gamma \cdot \left(\max_{0 \leq j \leq \ell} \alpha_j \cdot \sum_{i=0}^{\lfloor k/\ell \rfloor} \alpha_\ell^i + \max\{1, \alpha_0\} \right), & \text{if } B \in \mathcal{P}(\mathbb{T}_k) \\ 0 & \text{if } B \notin \mathcal{P}(\mathbb{T}_k). \end{cases}$$

We defer the proof of Theorem 21 to Appendix C. The block factorization of variance directly implies the relaxation time of the heat-bath block dynamics. We have the following corollary.

Corollary 22. *Under the same conditions as Theorem 21, if $\alpha_\ell < 1$ then the relaxation time of the heat-bath $\mathcal{P}(\mathbb{T}_k)$-dynamics is $O_{\alpha,\beta,\gamma}(n)$ on $\mathbb{T}_k$, the Δ-regular tree of depth k.*

We now give an overview on proving Theorem 2 which gives an optimal relaxation time bound for the neighboring edge dynamics on trees of maximum degree when $q \geq \Delta + 1$. Similar to the proof of Theorem 1, we will prove the $O_{\Delta,q}(n)$ relaxation time bound on $\mathbb{T}_k$ first. The crucial part is to verify the conditions in Theorem 21.

Lemma 23. *Let $\Delta \geq 2$, $q = \Delta + 1$. There exists $\ell = O(\Delta^2 \log^2 \Delta)$ such that the constants $\alpha = (\alpha_0, \cdots, \alpha_\ell)$, β, and γ in Theorem 21 exist with $\alpha_\ell = \frac{1}{2}$, $\alpha_j = q^{\Delta^{O(\ell)}}$ for all $j < \ell$, and $\beta = \gamma = q^{\Delta^{O(\ell)}}$.*

The proof of Lemma 23 is very similar to the proof of Lemma 13 for the Glauber dynamics. Following the high-level plan in Section 3.2 and Section 5, one can prove Lemma 23 by slightly modifying the construction of the canonical path. We omit the detailed proof for Lemma 23. Instead, we will give an proof overview for Lemma 23 in Section 6, where we will discuss why the edge dynamics can bypass the obstacles on Glauber dynamics in the canonical path analysis.

To obtain relaxation time bound for arbitrary trees, we also need a comparison argument for the heat-bath neighboring edge dynamics which is similar to Lemma 14.

Lemma 24. *Let $\mathbb{T} = (V, E)$ be a tree with root r and maximum degree Δ. Let $\mathbb{T}_H = (V_H, H \subseteq E)$ be a connected sub-tree of $\mathbb{T}$ containing r Let $q \geq \Delta + 1$, and μ, $\nu = \mu_H$ be the distribution of uniform q-edge-coloring on $\mathbb{T}$ and $\mathbb{T}_H$, respectively. If μ satisfies approximate edge-tensorization of variance with constant C, then ν satisfies approximate edge-tensorization of variance with constant $C(q+1)^2$.*

The proof of Lemma 24 is given in Appendix E.

Proof of Theorem 2. The proof follows from a similar argument as the proof of Theorem 1.

Given a tree $\mathbb{T}$ with maximum degree Δ and $q \geq \Delta + 1$. Let ν be the distribution of uniform q-edge-coloring on $\mathbb{T}$. We consider the distribution of uniform q-edge-coloring μ on the $(q-1)$-regular tree $\mathbb{T}_k$ for sufficiently large depth k and degree $\Delta' = q - 1$ such that $\mathbb{T}$ is a subtree of $\mathbb{T}_k$. Since $q \geq \Delta + 1$, we have $\Delta' \geq \Delta$ and such $\mathbb{T}_k$ must exist.

By Corollary 22 and Lemma 23, we know that when $q = \Delta' + 1$, the relaxation time of the edge dynamics on μ is $O_{\Delta',q}(n) = O_q(n)$. This gives $O_q(1)$ approximate edge-tensorization of variance of μ. Then, after applying Lemma 24, we know that ν also has $O_q(1)$ edge-approximate tensorization of variance. This proves the $O_{\Delta,q}(n)$ relaxation time for the edge-dynamics on ν. □

4 Approx. tensorization via root-tensorization: Proof of Theorem 11

In this section, we prove Theorem 11; this is the main inductive proof of approximate tensorization of variance for an arbitrarily large Δ-regular complete tree using the root-tensorization property for a small tree of height ℓ. Without loss of generality, we focus on μ_k. The approximate tensorization of variance for $\mu_k^{\star,1}$ follows from the same proof. Recall the definition of $\mathbb{T}_k, \mathbb{T}_k^\star$ and $\mu_k, \mu_k^{\star,1}$ in Definitions 9 and 10.

Fix parameters ℓ, $\boldsymbol{\alpha}$, and γ in the theorem. Our proof is based on an induction on the depth k of the regular tree $\mathbb{T}_k$. For every $k \geq 1$, we will show that μ satisfies approximate tensorization of variance with constants $C(e) = F_k(t)$ for $1 \leq t \leq k$ and $e \in L_t(\mathbb{T}_k)$. In other words, let $\mu = \mu_k$, we will prove the following inequality for $f : \Omega(\mu) \to \mathbb{R}$,

$$\mathbf{Var}_\mu[f] \leq \sum_{t=1}^{k} F_k(t) \sum_{e \in L_t(\mathbb{T}_k)} \mu[\mathbf{Var}_e[f]]. \tag{19}$$

The function $F.(\cdot)$ is defined as follows: for any $1 \leq t \leq s \leq \ell$, $F_s(t) = \gamma$; for any $s > \ell$,

$$F_s(t) = \begin{cases} F_{s-\ell}(t) & \text{if } 1 \leq t \leq s - \ell - 1 \\ \alpha_0 F_{s-\ell}(s-\ell) & \text{if } t = s - \ell \\ \alpha_{t-s+\ell} F_{s-\ell}(s-\ell) + \gamma & \text{if } t \geq s - \ell + 1. \end{cases} \tag{20}$$

Then, the proof of Theorem 11 is reduced to giving an upper bound for $F.(\cdot)$. We defer that part to Section 4.3.

4.1 Basic properties of variance

We begin with some basic properties about variance that will be useful in the proof of Theorem 11. The following lemma is a basic decomposition property of variance (e.g. see [MSW04, CLV21, CP21, Cap23]).

Lemma 25. *Let π be a distribution over $\Lambda \subset [q]^n$ and $f : \Omega \to \mathbb{R}$ be a function. For any subset $S \subseteq [n]$,*

$$\mathbf{Var}_\pi [f] = \pi[\mathbf{Var}_S [f]] + \mathbf{Var}_\pi [\pi_S[f]]. \tag{21}$$

Furthermore, if $\pi = \pi_{S_1} \times \pi_{S_2} \times \ldots \times \pi_{S_\ell}$ is a product distribution of ℓ marginal distributions, where $S_1, S_2, \ldots, S_\ell$ is a partition of $[n]$, then

$$\mathbf{Var}_\pi [f] \leq \sum_{i=1}^{\ell} \pi[\mathbf{Var}_{S_i} [f]]. \tag{22}$$

The identity in (21) is called the law of total variance. The inequality in (22) states that approximate tensorization of variance holds for product distributions with constant $C = 1$ which is the optimal constant.

In this paper, the distribution μ is over list edge colorings of a graph. Let $G = (V, E)$ be a graph. Let μ over $[q]^E$ be a distribution satisfying the following conditional independence property. For three disjoint $A, B, C \subseteq E$ such that the removal of C disconnects A and B in the line graph of G,

$$\forall \sigma \in \Omega(\mu_C), \quad \mu_{A \cup B}^\sigma = \mu_A^\sigma \times \mu_B^\sigma. \tag{23}$$

For distributions with the conditional independence property (e.g., the uniform distribution of (list) edge colorings), the following lemma holds.

Lemma 26 ([MSW04]). *Let $G = (V, E)$ be a graph and μ be distribution over $[q]^E$ satisfying the conditional independence property in (23). For any $f : \Omega \to \mathbb{R}$, any $S_1, S_2 \subseteq E$ such that $\partial S_1 \cap S_2 = \varnothing$,*

$$\mu[\mathbf{Var}_{S_1} [\mu_{S_2}[f]]] \leq \mu[\mathbf{Var}_{S_1} [\mu_{S_1 \cap S_2}[f]]], \tag{24}$$

where $\partial S_1 := \{e \in E \setminus S_1 \mid \exists h \in S_1, e \cap h \neq \varnothing\}$ denotes the exterior boundary of S_1.

The lemma in [MSW04] is stated for distributions defined over $[q]^V$ instead of $[q]^E$ and the proof is omitted there, and hence we provide a proof of Lemma 26 in Appendix F for completeness.

4.2 Establishing approximate tensorization (19)

Under the condition of Theorem 11, for every $k \geq 1$, we do the induction as follows.

Base case ($k \le \ell$): When $k \le \ell$, (19) follows from the second assumption of Theorem 11 with $F_k(t) = \gamma$.

Inductive case ($k > \ell$): Suppose (19) holds for all smaller $k' < k$. We prove that it also holds for k. For convenience, let $\mathbb{T} = \mathbb{T}_k$. Let $R = \cup_{i=k-\ell+1}^{k} L_i(\mathbb{T})$ be the edge set of the bottom ℓ levels of $\mathbb{T}$. According to the law of total variance (21), it holds that

$$\mathbf{Var}_\mu[f] = \mu[\mathbf{Var}_R[f]] + \mathbf{Var}_\mu[\mu_R[f]]. \tag{25}$$

We give the following bound for the first term $\mu[\mathbf{Var}_R[f]]$ in the RHS of (25).

Claim 27. *Assuming Item 2 of Theorem 11, we have*

$$\mu[\mathbf{Var}_R[f]] \le \gamma \sum_{i=1}^{\ell} \sum_{h \in L_{k-\ell+i}(\mathbb{T})} \mu[\mathbf{Var}_h[f]]. \tag{26}$$

We now prove the theorem assuming the above claim. The proof of Claim 27 will be given later.

Next, we bound the second term $\mathbf{Var}_\mu[\mu_R[f]]$ in the RHS of (25) using the inductive hypothesis. Note that for any $\sigma \in \Omega(\mu)$, the value of the function $\mu_R[f](\sigma)$ depends only on $\sigma_{\mathbb{T}\setminus R}$. By viewing $\mu_R[f]$ as a function $f' : \Omega(\mu_{\mathbb{T}\setminus R}) \to \mathbb{R}$, we have $\mathbf{Var}_\mu[\mu_R[f]] = \mathbf{Var}_{\mu_{\mathbb{T}-R}}[f'] = \mathbf{Var}_{\mu_{k-\ell}}[f']$, where the last equation uses the fact that the marginal distribution $\mu_{\mathbb{T}\setminus R}$ is the same as $\mu_{k-\ell}$, and $\mu_{k-\ell}$ is the uniform distribution over proper q-edge-coloring in the tree $\mathbb{T}_{k-\ell}$. For convenience, we use $\mathbb{T}^h$ to denote the subtree with the hanging root edge h and use $\widetilde{\mathbb{T}}^h := \mathbb{T}^h \setminus h$ to denote the subtree beneath the edge h. By the inductive hypothesis on $\mu_{k-\ell} = \mu_{\mathbb{T}\setminus R}$, we have

$$\mathbf{Var}_\mu[\mu_R[f]] \le \sum_{t=1}^{k-\ell} F_{k-\ell}(t) \sum_{h \in L_t(\mathbb{T})} \mu_{\mathbb{T}\setminus R}[\mathbf{Var}_h[f']]$$

$$= \sum_{t=1}^{k-\ell-1} F_{k-\ell}(t) \sum_{h \in L_t(\mathbb{T})} \mu[\mathbf{Var}_h[\mu_R[f]]] + F_{k-\ell}(k-\ell) \sum_{h \in L_{k-\ell}(\mathbb{T})} \mu[\mathbf{Var}_{\mathbb{T}^h}[\mu_R[f]]] \tag{27}$$

$$\le \sum_{t=1}^{k-\ell-1} F_{k-\ell}(t) \sum_{h \in L_t(\mathbb{T})} \mu[\mathbf{Var}_h[f]] + F_{k-\ell}(k-\ell) \sum_{h \in L_{k-\ell}(\mathbb{T})} \mu[\mathbf{Var}_{\mathbb{T}^h}[\mu_{\widetilde{\mathbb{T}}^h}[f]]], \tag{28}$$

where (28) holds by (24); and (27) holds because of the following:

- for all $h \in L_t(\mathbb{T})$ with $t < k - \ell$, all the neighboring edges of h are in $\mathbb{T} \setminus R$, i.e. $\partial h \subseteq \mathbb{T} \setminus R$, by conditional independence, $\mu_{\mathbb{T}\setminus R}[\mathbf{Var}_h[f']] = \mu[\mathbf{Var}_h[\mu_R[f]]]$;

- for all $h \in L_{k-\ell}(\mathbb{T})$, $\mu_{\mathbb{T}\setminus R}[\mathbf{Var}_h[f']] = \mu[\mathbf{Var}_{\mathbb{T}^h}[\mu_R[f]]]$ because for any $\sigma \in \Omega(\mu)$, the value of $\mu_R[f](\sigma)$ is independent from $\sigma_{\mathbb{T}_h\setminus h}$.

Since $\mu[\mathbf{Var}_{\mathbb{T}^h}[\mu_{\widetilde{\mathbb{T}}^h}[f]]] = \sum_{\sigma \in \Omega(\mu_{\mathbb{T}\setminus\widetilde{\mathbb{T}}^h})} \mu_{\mathbb{T}\setminus\widetilde{\mathbb{T}}^h}(\sigma)\mathbf{Var}_{\mu^\sigma}[\mu_{\widetilde{\mathbb{T}}^h}[f]]$. Fix one σ. One can view f as a function from $\Omega(\mu^\sigma_{\widetilde{\mathbb{T}}^h})$ to $\mathbb{R}$. By applying Item 1 of Theorem 11 to each σ individually and rearranging, it holds that

$$\sum_{h \in L_{k-\ell}(\mathbb{T})} \mu[\mathbf{Var}_{\mathbb{T}^h}[\mu_{\widetilde{\mathbb{T}}^h}[f]]] \le \sum_{i=0}^{\ell} \alpha_i \sum_{g \in L_{k-\ell+i}(\mathbb{T})} \mu[\mathbf{Var}_g[f]]. \tag{29}$$

Combining (25), (26), (28), and (29), it holds that

$$\mathbf{Var}_\mu [f] \leq \sum_{t=1}^{k-\ell-1} F_{k-\ell}(t) \sum_{h \in L_t(\mathbb{T})} \mu[\mathbf{Var}_h [f]]$$
$$+ F_{k-\ell}(k-\ell) \cdot \alpha_0 \sum_{h \in L_{k-\ell}(\mathbb{T})} \mu[\mathbf{Var}_h [f]]$$
$$+ \sum_{i=1}^{\ell} (\alpha_i F_{k-\ell}(k-\ell) + \gamma) \sum_{h \in L_{k-\ell+i}(\mathbb{T})} \mu[\mathbf{Var}_h [f]].$$

The above inequality immediately gives the recurrence of F.

Proof of Claim 27. Note that R is a set of disjoint trees $\mathbb{T}^{(j)}$ with depth ℓ for $j \in [z]$. We use $\mathbb{T}^{(j,\star)}$ to denote the tree obtained by combining $\mathbb{T}^{(j)}$ with the edge connecting the root of $\mathbb{T}^{(j)}$ to its parent in $\mathbb{T}$. Conditional on any coloring $\sigma \in \Omega(\mu_{E \setminus R})$, μ^σ is a product distribution over marginal distributions of all $\mathbb{T}^{(j,\star)}$'s and $H = E \setminus (\cup_{i=1}^z \mathbb{T}^{(j,\star)})$. We have

$$\mu[\mathbf{Var}_R [f]] = \sum_{\sigma \in \Omega(\mu_{E \setminus R})} \mu_{E \setminus R}(\sigma) \mathbf{Var}_{\mu^\sigma} [f]$$
$$\leq \sum_{\sigma \in \Omega(\mu_{E \setminus R})} \mu_{E \setminus R}(\sigma) \sum_{j=1}^z \mu^\sigma [\mathbf{Var}_{\mathbb{T}^{(j,\star)}} [f]] = \sum_{j=1}^z \mu[\mathbf{Var}_{\mathbb{T}^{(j,\star)}} [f]], \qquad (30)$$

where the inequality (30) holds because of (22) and $\mu^\sigma[\mathbf{Var}_H [f]] = 0$. Note that conditioning on any feasible coloring σ of $E \setminus \mathbb{T}^{(j)}$, $\mu^\sigma_{T^{(j,\star)}}$ is the same as the distribution $\mu_\ell^{\star,1}$ in Item 2 of Theorem 11 by the symmetry of colors. We can open the expectation in $\mu[\mathbf{Var}_{\mathbb{T}^{(j,\star)}} [f]]$ then apply Item 2 of Theorem 11 to obtain

$$\mu[\mathbf{Var}_{\mathbb{T}^{(j,\star)}} [f]] = \sum_{\sigma \in \Omega(\mu_{E \setminus \mathbb{T}^{(j,\star)}})} \mu_{E \setminus \mathbb{T}^{(j,\star)}}(\sigma) \mathbf{Var}_{\mu^\sigma} [f]$$
$$\leq \sum_{\sigma \in \Omega(\mu_{E \setminus \mathbb{T}^{(j,\star)}})} \mu_{E \setminus \mathbb{T}^{(j,\star)}}(\sigma) \gamma \sum_{i=1}^{\ell} \sum_{h \in L_i(\mathbb{T}_\ell^\star)} \mu^\sigma [\mathbf{Var}_h [f]]$$
$$\leq \gamma \sum_{i=1}^{\ell} \sum_{h \in L_i(\mathbb{T}^{(j,\star)})} \mu[\mathbf{Var}_h [f]]. \qquad (31)$$

By the definition of $\mathbb{T}^{(j)}$, we know that $h \in L_i(\mathbb{T}^{(j)})$ means $h \in L_{k-\ell+i}(\mathbb{T})$. So, Claim 27 is proved by combining (30) and (31). $\qquad \square$

4.3 Upper bound for the recursion F

In this section, we will give an upper bound for the recursion $F.(\cdot)$ defined in (20) and then finish the proof of Theorem 11 for μ_k. The proof of Theorem 11 for $\mu_k^{\star,1}$ follows from the same proof.

Let $S_k := \{a \in \mathbb{Z} \mid a \equiv k \pmod{\ell}\}$ be a set of integer. By resolving the recursion in (20), we claim that $F_k : [k] \to \mathbb{R}$ has the following upper bound:

$$\forall 1 \leq t \leq k, \quad \widehat{F}_k(t) := \gamma \cdot \left(\left(\sum_{i=0}^{|[t-1] \cap S_k|} \alpha_\ell^i \right) \cdot \alpha_{t-\max([t-1] \cap S_k)} + 1 \right) \cdot \alpha_0^{\mathbf{1}[t \in S_k \text{ and } t \neq k]}. \qquad (32)$$

18

Proof of Theorem 11. The upper bound stated in Theorem 11 follows directly from (32). □

Now, we only left to prove (32). We will finish the proof by induction. For the base case, that $1 \leq t \leq k \leq \ell$, it holds that $F_k(t) = \gamma \leq \gamma(1 + \alpha_t) = \widehat{F}_k(t)$. For the inductive case where $k \geq \ell$, suppose for any $s < k$ and $t \leq s$, it holds that $F_s(t) \leq \widehat{F}_s(t)$. By definition, $S_k = S_{k-\ell}$. This implies $F_k(t) = F_{k-\ell}(t) \leq \widehat{F}_{k-\ell}(t) = \widehat{F}_k(t)$ for $t \leq k - \ell - 1$ and $F_k(k - \ell) = F_{k-\ell}(k - \ell) \cdot \alpha_0 \leq \widehat{F}_{k-\ell}(k - \ell) \cdot \alpha_0 = \widehat{F}_k(k - \ell)$ by definition. For $t \geq k - \ell + 1$, let $j := t - (k - \ell)$, we have

$$F_k(t) = \alpha_j F_{k-\ell}(k - \ell) + \gamma \leq \alpha_j \widehat{F}_{k-\ell}(k - \ell) + \gamma$$
$$= \gamma \left(\alpha_j \left(\left(\sum_{i=0}^{|[k-\ell-1] \cap S_k|} \alpha_\ell^i \right) \cdot \alpha_\ell + 1 \right) + 1 \right)$$
$$= \gamma \left(\alpha_j \left(\sum_{i=0}^{|[k-\ell] \cap S_k|} \alpha_\ell^i \right) + 1 \right) = \widehat{F}_k(t).$$

This finishes the proof of (32).

5 Coupling and canonical paths

In this section, we give a proof for Lemma 19, which shows that for the Δ-regular complete tree of height $\ell = O(\Delta^2 \log^2 \Delta)$ there is a set of couplings $\mathcal{C}$ and a set of canonical paths Γ with good congestion on the leaf edges. This lemma then completes the proof of Lemma 13 as presented in Section 3.2.

In Section 5.1, we will describe the construction of canonical paths. Then, we analyze the congestion for these canonical paths and finish the proof of Lemma 19 in Section 5.3 by bounding the number of canonical paths that may go through a specific coloring γ via leaf edges (see Lemma 31); and the probability of such a coloring γ appearing (see Lemma 32). Then, we finish the proofs of Lemmas 31 and 32 in Sections 5.3 and 5.4, respectively.

5.1 Construction

Fix $a, b \in [q - d]$ with $a \neq b$, we construct a coupling $\mathcal{C}_{ab}$ between μ^{ra} and μ^{rb}. Consider a random sample $\sigma \sim \Omega(\mu^{ra})$. The root hanging edge r must receive the color $\sigma_r = a$. Starting from r, we can find the maximal (a, b)-alternating path $e_0, e_1, e_2, \ldots, e_s$, which we denote as $\mathcal{E}^*$. Note that $\mathcal{E}^* = (e_0, e_1, e_2, \ldots, e_s)$ has the following properties: $e_0 = r$; $\sigma(e_i) = a$ for even i and $\sigma(e_j) = b$ for odd j. We can "flip" the colors a and b in this alternating path $\mathcal{E}^*$ to obtain a new coloring τ (by flip we mean interchange the colors a and b). We denote this flip operation by $\tau := \sigma \oslash_r b$. This notation means, given a coloring σ, an edge r, and a color b, we find the maximal $(\sigma(r), b)$-alternating path containing r, and flip the colors $(\sigma(r), b)$ in the maximal alternating path to obtain the new coloring τ. The following observation is easy to verify.

Observation 28. *For $\sigma \sim \mu^{ra}$, let $\tau = \sigma \oslash_r b$. Then, $\tau \sim \mu^{rb}$ and τ is uniquely determined by σ and b.*

Hence, this construction of (σ, τ) defines a coupling $\mathcal{C}_{ab}$ of μ^{ra} with μ^{rb}; we refer to this coupling $\mathcal{C}_{ab}$ as the *flip coupling*. Note the coupling is symmetric, i.e., $\mathcal{C}_{ab} = \mathcal{C}_{ba}$. Let $\mathcal{C} = \cup_{a,b} \mathcal{C}_{ab}$.

Fix an initial coloring $\sigma \in \Omega(\mu^{ra})$ and the final coloring $\tau = \tau(\sigma, b) \in \Omega(\mu^{rb})$ where $\tau = \sigma \oslash_r b$. Let $v_0, v_1, \ldots, v_{s+1}$ list the vertices of $\mathcal{E}^*$ in order from the root downwards, and let $\mathcal{E}^* = (e_0, e_1, \ldots, e_s)$ be the edges in downwards order. Therefore, $r = e_0 = \{v_0, v_1\}$ and $e_i = \{v_i, v_{i+1}\}$.

We now detail the construction of the canonical path $\gamma^{\sigma,\tau}$ from σ to τ. Recall, $\mathcal{E}^* = (e_0, e_1, e_2, \ldots, e_s)$ is the (a,b)-alternating path in σ which we flip to obtain τ and the path $\gamma^{\sigma,\tau}$ needs to perform this flip by a sequence of single-edge updates (i.e., Glauber transitions).

Fix an arbitrary ordering $\mathcal{O}$ on all colors in $[q]$ such that the colors a and b are the last two colors. The following is the high level idea for constructing the canonical path $\gamma^{\sigma,\tau}$ from σ to τ, see Figure 1 for an illustration.

Stage-I: For every $e_i = \{v_i, v_{i+1}\} \in \sigma \oplus \tau$, where $i \in [s] = \{1, 2, \ldots, s\}$ and i is odd, we would like to change the color $\sigma(e_i) = b$ to some color $c_i \notin \{a, b\}$. If e_i has multiple such available colors (i.e., do not appear in the neighborhood of e_i in σ) that do not in $\{a, b\}$ then we let c_i be the first such available color in the ordering $\mathcal{O}$. If e_i does not have any available color, excluding its current color b (note that a is not available as $\sigma(e_{i-1}) = a$) then we will define a path $\mathcal{E}_i$ from v_i downwards using the ordering $\mathcal{O}$; this path $\mathcal{E}_i$ may reach a leaf edge. The construction of this path $\mathcal{E}_i$ will be detailed later. We let c_i be the initial color in σ of the first edge in $\mathcal{E}_i$, and after recoloring $\mathcal{E}_i$ then color c_i will be available for e_i. The new color for each edge in $\mathcal{E}_i$ will be the initial color in σ of its descendant edge in the path $\mathcal{E}_i$ (and we will recolor these edges from the leaves upwards).

We recolor all of these paths $\mathcal{E}_i$ and the odd edges of $\mathcal{E}^*$ in the following order. Let $\mathcal{E} = (\bigcup_{odd\ i} e_i) \cup (\bigcup_i \mathcal{E}_i)$ denote the set of edges we are recoloring in this stage. We order the edges in $\mathcal{E}$ so that edges in lower levels (closer to the leaves) are before higher levels, and thus the leaf edges are first. Moreover, within a level all of the edges in $\cup_i \mathcal{E}_i$ are before those in $\mathcal{E}^*$; this ensures that the edges in $\mathcal{E} \cap \mathcal{E}^*$ can be recolored to c_i. To change the color of e_i, this process may change all colors along a path $(e_i \cup \mathcal{E}_i)$, where $(e_i \cup \mathcal{E}_i)$ intersects with $\sigma \oplus \tau$ only at edge e_i.

We describe how to construct the path $\mathcal{E}_i$ for odd $i \in [s]$. If $e_i = \{v_i, v_{i+1}\}$ has an available color $c_i \notin \{a, b\}$, then let $\mathcal{E}_i = \varnothing$. Otherwise, let $N_e(v_{i+1})$ denote the set of edges incident to v_{i+1} and let c_i be the first color in $[q] \setminus \sigma(N_e(v_{i+1}))$ according to the ordering $\mathcal{O}$. We call c_i the first available color for vertex v_{i+1} in σ. Note that the color $c_i \notin \{a, b\}$ because $|[q] \setminus \sigma(N_e(v_{i+1}))| = q - \Delta = 2$ and the color b must in $\sigma(N_e(v_{i+1}))$. The vertex v_{i+1} must have an available color that is not a nor b and that color would be the first one according to $\mathcal{O}$. We find the incident edge $e = \{v_i, v\}$ to v_i whose color is c_i in σ. Such edge e must exist, otherwise $c_i \notin \{a, b\}$ is an available color for e_i and we do not need to construct $\mathcal{E}_i$. Hence, we defined the first edge e in the path $\mathcal{E}_i$. All other edges can be constructed by the following recursive procedure. Specifically, suppose $\{u_j, u_{j+1}\}$ is the j-th edge in $\mathcal{E}_i$ and u_{j+1} is a child of u_j. If $\{u_j, u_{j+1}\}$ has two available colors, we stop the construction of $\mathcal{E}_i$. Otherwise, let c be the first available color for u_j in σ; we find the edge e' incident to u_{j+1} whose color is c in σ; and let e' be the $(j+1)$-th edge in $\mathcal{E}_i$. This process must stop once it touches a leaf edge but it may stop earlier before going down to the leaves.

We now detail the process of changing colors along the paths $(e_i \cup \mathcal{E}_i)$ for odd $i \in [s]$. For two different paths $(e_i \cup \mathcal{E}_i)$ and $(e_{i'} \cup \mathcal{E}_{i'})$, their distance is at least 2 in the line graph. Hence, changing the colors in $(e_i \cup \mathcal{E}_i)$ has no effect on the process of changing colors in $(e_{i'} \cup \mathcal{E}_{i'})$. For a specific path $e_i \cup \mathcal{E}_i = \{e_i =: h_0, h_1, h_2, \ldots, h_t\}$, we first change the color of h_t to the first available color c for h_t in σ with $c \neq \sigma(h_t)$. For each j from $t-1$ down to 0, after changing the color of h_{j+1}, $\sigma(h_{j+1})$ becomes an available color for h_j and we change the color of h_j to $\sigma(h_{j+1})$.

After Stage-I, for any even $j \in \{0\} \cup [s]$, we claim that color b is available for edge $e_j = \{v_j, v_{j+1}\}$ in the current coloring. For even j, both e_{j-1} and e_{j+1} switched to colors not in $\{a, b\}$. Note that e_j may have an incident edge e in $\mathcal{E}_{j+1}$ such that $e \cap e_j = \{v_{j+1}\}$. We need to show that the edge e cannot take b in the current coloring. This is because, in the initial coloring σ, v_{j+1} has incident edges with colors a and b. Note that after Stage-I, e either takes the first available color for v_{j+1} in σ ($\mathcal{E}_i$ has at least 2 edges) or an available color of e in σ (e is the only edge in $\mathcal{E}_i$). In both cases, the new color cannot be b and cannot be a.

A detailed illustration of **Stage-I** is given at Figure 2.

20

Stage-II: For every $e_i = \{v_i, v_{i+1}\} \in \sigma \oplus \tau$, where $i \in \{0\} \cup [s]$ and i is even, we change the color of e_i from $\sigma(e_i) = a$ to color b.

Stage-III: In this stage we recolor all of $\mathcal{E}_i$, for all odd i, and the odd edges of $\mathcal{E}^*$ by the following process. First, consider the same set $\mathcal{E}$ from Stage-I. Then, order the edges in $\mathcal{E}$ in reverse order from Stage-I's ordering, and hence the edges closer to the root are earlier in the ordering and edges in $\mathcal{E}^*$ are before $\mathcal{E}_i$. For every edge $e \in \mathcal{E}$ we recolor it to $\tau(e)$ (proceeding in the order we just defined). Note that for $e \in \mathcal{E}^* \cap \mathcal{E}$ then $\tau(e) = a$ and for $e \in \mathcal{E}_i$ then $\tau(e) = \sigma(e)$ is its original color in σ.
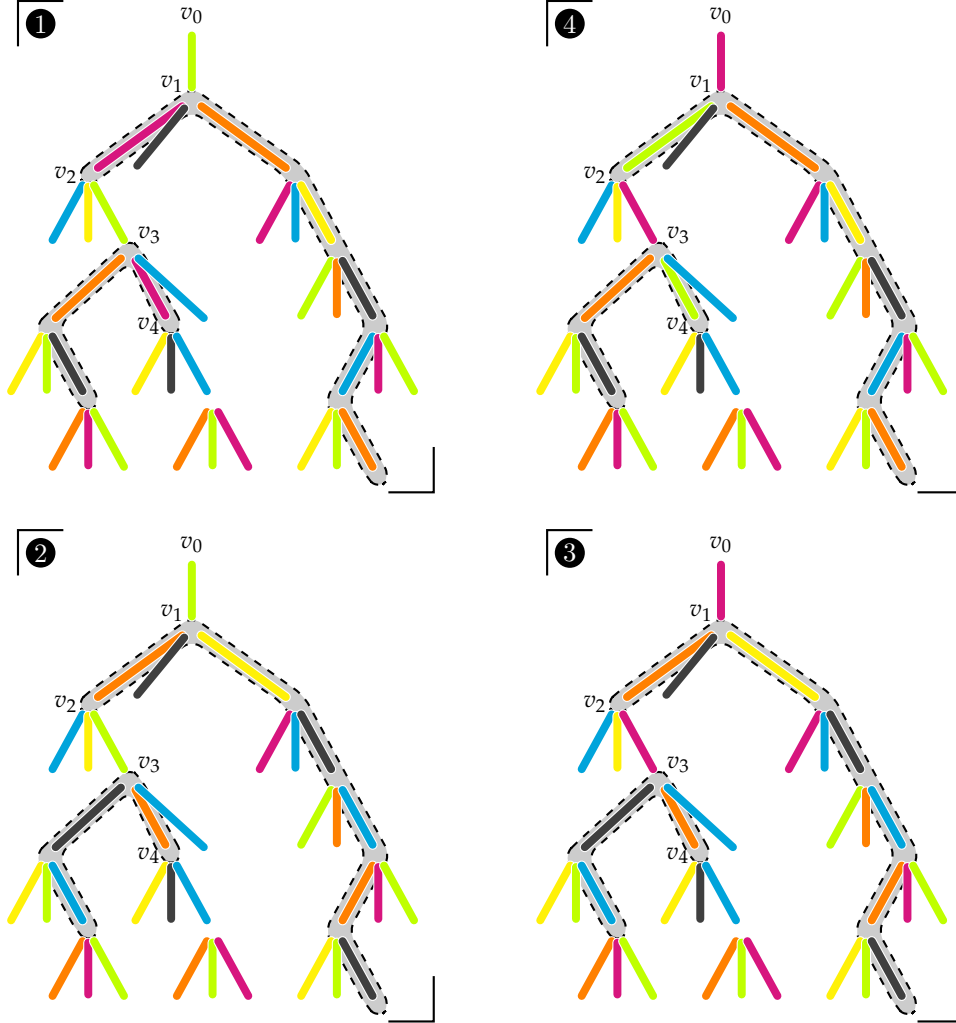

Figure 1: The figure illustrates the canonical path $\gamma^{\sigma,\tau}$. For simplicity, we only draw edges whose color will be used in the construction rather than the entire tree. The edge-coloring in ❶ is σ and the coloring in ❹ is τ. Let $a = \bullet$ and $b = \bullet$, we mark the vertices in (a,b)-alternating path $\mathcal{E}^*$ as $(v_0, v_1, v_2, v_3, v_4)$. Then, we fix the order $\mathcal{O}$ of colors as $\mathcal{O} = \{\bullet \prec \bullet \prec \bullet \prec \bullet \prec \bullet \prec \bullet\}$; note, the colors $\bullet, \bullet$ are the last two colors in this order. The odd edges $e_i \in \mathcal{E}^*$ and their associated paths $\mathcal{E}_i$ are marked by dashed lines (and grey background). The coloring in ❷ is obtained after Stage-I, and hence after we recolored the edges in $\mathcal{E}$ in order from the leaves up. Then the coloring in ❸ is after Stage-II, and finally the coloring in ❹ is after Stage-III and is τ.
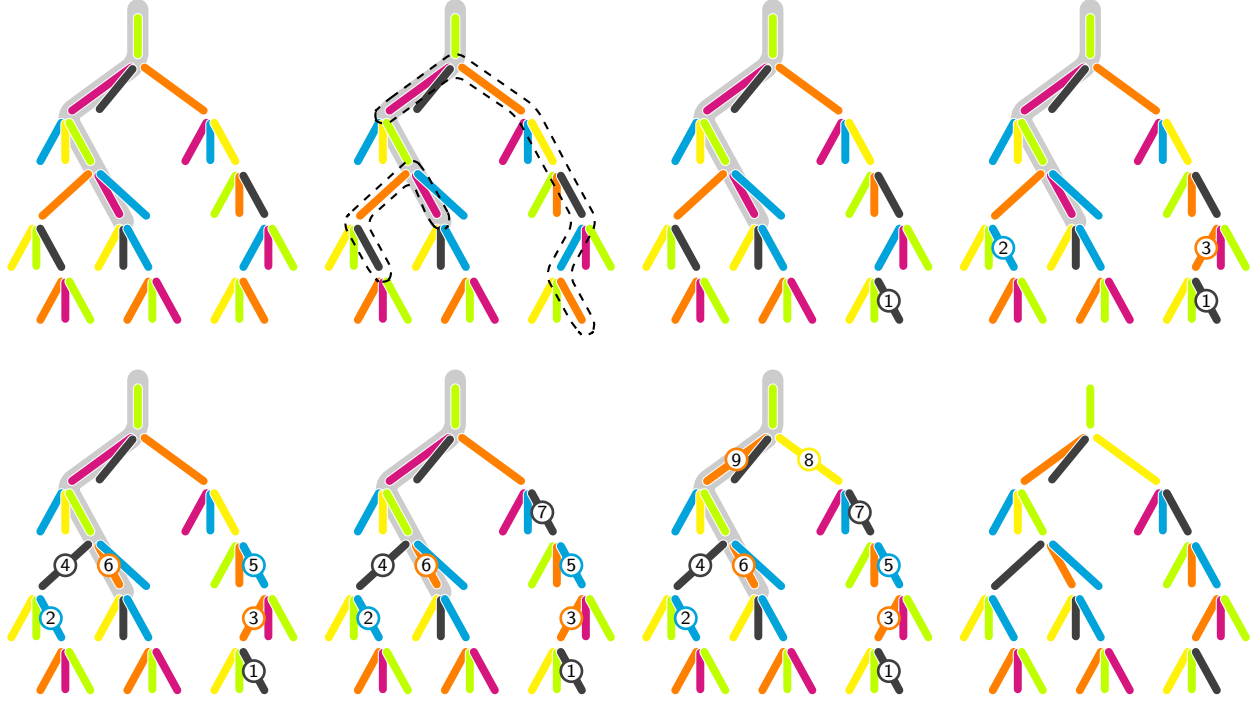
Figure 2: The figure details **Stage-I** of the canonical path. For convenience, we call the sub-figure located in the i-th row and j-th column as fig(i, j). The canonical path is as follows. First, as in fig$(1, 1)$, we find the ($\bullet$, $\bullet$)-alternating path; this is marked with a shadowed background. Then, for the odd edges on the alternating path (i.e., the $\bullet$-edges), we want to change their color to some color other than $\bullet$ and $\bullet$. To do that, we run Stage-I according to the priority of colors $\bullet \prec \bullet \prec \bullet \prec \bullet \prec \bullet \prec \bullet$ (see also Figure 1) and get two lists of edges. We mark these lists $\mathcal{E}$ with dashed circles in fig$(1, 2)$. Intuitively, to change the color of the odd edges on the alternating path, we need to change the color of edges in these lists. Then, as showed from fig$(1, 3)$ to fig$(2, 3)$, we change the color of the edges in the lists according to the order of their depth. For the edges on the same level, to break the tie, we first consider edges from left to right, but with an exception that we keep the edge in the alternating path to be the last one. We add number on these edges to denote this order. Lastly, when we change the color of some edge i, we always change its color to a different color which has the lowest priority among all the available colors.

Proposition 29. *If $q = \Delta + 2$, the above process constructs a feasible canonical path $\gamma^{\sigma, \tau}$ from σ to τ.*

Proof. It is easy to verify the canonical path constructed at Stage-I and Stage-II is a feasible path starting from σ to γ, where γ denotes the coloring obtained after Stage-II. Stage-III constructs a canonical path from γ to τ. We only need to show that any coloring in the canonical path constructed at Stage-III is a feasible edge coloring.

Consider the following process $\mathcal{P}$, we run Stage-I starting from the coloring τ and obtain a canonical path from τ to a coloring γ'. Specifically, we first find the maximal (b, a)-alternating path $\mathcal{E}^*$ from r in τ. We then do the same procedure described in Stage-I except that we flip the roles of σ and τ, and also the roles of a and b.

Claim 30. *The canonical path generated by $\mathcal{P}$ is the reverse of the canonical path generated by Stage-III. In particular, $\gamma = \gamma'$.*

22

It is easy to verify Stage-I generates a feasible canonical path, the above claim implies the canonical path generated by Stage-III is also feasible, which proves the proposition.

We now prove Claim 30. We use Stage-I(τ) to denote the process $\mathcal{P}$, where the notation emphasizes that $\mathcal{P}$ is the Stage-I starting from τ. Similarly, we use Stage-I(σ) to denote the original process starting from σ (described in Section 5.1). To prove the claim, we show that two processes Stage-I(σ) and Stage-I(τ) find the same set of edges $\mathcal{E}$, and for any $e \in \mathcal{E}$, two processes change the color of e to the same color $\gamma(e) = \gamma'(e)$.

We first verify that Stage-I(τ) generates the same paths $\mathcal{E}^*$ and $(\mathcal{E}_i)_{i \in [s] \text{ is odd}}$ as Stage-I(σ). It is easy to see that the alternating path $\mathcal{E}^*$ is the same. If a path $\mathcal{E}_i$ is empty in Stage-I(σ), then e_i must have an available color in σ that is not in $\{a, b\}$, the same color must be available for e_i in τ, and the path $\mathcal{E}_i$ is also empty in Stage-I(τ). If a path $\mathcal{E}_i$ is not empty in Stage-I, we need to show that $\mathcal{E}_i$ is also constructed by Stage-I(τ). The first edge of $\mathcal{E}_i$ is determined by the first available color of v_{i+1}. Since a, b are the last two colors in $\mathcal{O}$ and $q = \Delta + 2$, in both Stage-I(τ) and Stage-I(τ), the first available color is not in $\{a, b\}$. Then, the two processes find the same first available color and the same first edge in $\mathcal{E}_i$. The path $\mathcal{E}_i$ intersects with $\mathcal{E}^*$ only at the vertex v_i. Since σ and τ agree everywhere except the alternating path and $\sigma(N_e(v_i)) = \tau(N_e(v_i))$, if the first edge is the same, it is easy to verify that all other edges are the same. Hence, the two processes obtain the same set of edges $\mathcal{E}$. We only need to show that for any $e \in \mathcal{E}$, Stage-I(τ) changes the color $\tau(e)$ to the color $\gamma(e)$ (namely, $\gamma'(e) = \gamma(e)$), which proves the claim.

For each odd i, e_i intersects with the first edge in $\mathcal{E}_i$ at the vertex v_i. We can put e_i before the first edge of $\mathcal{E}_i$ to obtain a path $\mathcal{E}_i^+$. Note that all $\mathcal{E}_i^+$ are not adjacent to each other. It suffices to analyze one path $\mathcal{E}_i^+$. Say $\mathcal{E}_i^+$ has edges $e_1, h_1, h_2, \dots, h_t$. Since σ and τ differ only at the alternating path, for all h_k with $k \geq 2$, h_k is not adjacent to $\mathcal{E}^*$, it is easy to see Stage-I(τ) and Stage-I(σ) change h_k to the same color, which implies $\gamma'(h_k) = \gamma(h_k)$. We will take care of h_1 and e_i in the following two paragraphs.

Consider the edge h_1. Suppose $\mathcal{E}_i$ only has edge h_1. In Stage-I(σ), the color of h_1 is changed to the first available color c of h_1 in σ such that $c \neq \sigma(h_1)$. Hence $\gamma(h_1) = c$. In Stage-I(τ), the color of h_1 is changed to the first available color c' of h_1 in τ such that $c' \neq \tau(h_1)$. Since h_1 contains the vertex v_i, it holds that $\sigma(N_e(v_i)) = \tau(N_e(v_i))$ and $\{a, b\} \subseteq \sigma(N_e(v_i))$. Since σ and τ take the same color for edges not in $\mathcal{E}^*$, One can verify that $c = c'$. Suppose $\mathcal{E}_i$ has at least 2 edges. In Stage-I(σ), the color of h_1 is changed to the first available color c of v_i in σ. In Stage-I(τ), the color of h_1 is changed to the first available color c' of v_i in τ. Again, by $\sigma(N_e(v_i)) = \tau(N_e(v_i))$, we know $c = c'$.

Consider the edge e_i. If e_i has an available color $c_i \notin \{a, b\}$ in σ, it must have the same available color in τ. In this case, e_i is changed to the same color in Stage-I(σ) and Stage-I(τ). If e_i does not have such an available color in σ, then it does not have it in τ. In this case, in Stage-I(σ), e_i is changed to the first available color c of v_{i+1} in σ. In Stage-I(τ), e_i is changed to the first available color c' of v_{i+1} in τ. Since, a and b are last two colors in the ordering $\mathcal{O}$, $c \notin \{a, b\}$ and $c' \notin \{a, b\}$. One can verify $c = c'$.

Combining all the above analysis together proves the claim, which proves the proposition. $\square$

5.2 Analysis of Congestion (Proof of Lemma 19)

In this section, we use the coupling and canonical path constructed above to prove Lemma 19. We will also specify the parameter ℓ for the lemma. We first point out that the bound for ξ_j with $j < \ell$ is trivial. By the definition of congestion,

$$\xi_j \leq \sum_{(\gamma, \gamma'): \gamma \oplus \gamma' \in L_j} \frac{1}{\mu(\gamma)^{\frac{1}{q}}} \leq q^{d^{O(\ell)}}$$

because the total number of list colors is at most $q^{d^{O(\ell)}}$. This bound holds for any choice of ℓ. Our main task is to bound the congestion $\xi_\ell = \max_{a,b} \xi_\ell^{ab} \le \frac{1}{2(\ell+1)}$ where

$$\xi_\ell^{ab} = \sum_{s=(\gamma,\gamma')\in\Lambda(\mu):\gamma\oplus\gamma'\in L_\ell} \frac{\left(\mathbf{Pr}_{(\sigma,\tau)\sim\mathcal{C}_{ab}}[s\in\gamma^{\sigma,\tau}]\right)^2}{\mu(\gamma)Q_\mu^{\text{Glauber}}(\gamma,\gamma')}, \tag{33}$$

where $\Lambda(\mu) \subseteq \Omega(\mu) \times \Omega(\mu)$ is the set of feasible transitions made by the Glauber dynamics.

Proof Overview: We first provide a high-level proof overview before diving into the details. Note that Stage-II only changes the colors of edges in $\mathcal{E}^*$ that are in even levels. If we set ℓ to be an odd number, then we only change the colors of leaf edges in Stage-I and Stage-III. Let us consider a move (γ, γ') in Stage-I such that $\gamma \oplus \gamma' \in L_\ell$. The analysis is similar if it is in Stage-III. To bound the congestion, the key step is to recover all possible initial colorings σ that can use the move (γ, γ') in Stage-I of the canonical path $\gamma^{\sigma,\tau}$, where the final coloring τ is fixed by σ. According to the ordering of edges in $\mathcal{E}$ defined in Stage-I, we always first change the colors of leaves, which implies for any non-leaf edges e, $\sigma(e) = \gamma(e)$. Now, we only need to recover the colors of leaf edges in σ. Furthermore, given γ, we can almost recover all the paths $\mathcal{E}^*$ and $(\mathcal{E}_i)_{i\in[s] \text{ is odd}}$ constructed in Stage-I. The only part we do not know is that if a path recovered by γ ends at level $\ell - 1$, then in σ, this path may end at level $\ell - 1$ or end at level ℓ. The second case may occur because when we move from σ to γ along the canonical path, we may change the colors of some leaf edges, which reduces the length of the path by 1. Hence, these paths can help us find all leaf edges where σ and γ may disagree. Suppose that in the paths constructed by γ, there are P paths ending at level $\ell - 1$. Then there are $(1 + (\Delta - 1)(q - \Delta))^P < (2\Delta)^P$ possible initial colorings σ. This is because for each of these paths, σ may have the same path (1 possibility) or have one additional leaf edge $(\Delta - 1$ possibilities) and there are at most $(q - \Delta) = 2$ ways the give the colors to the additional leaf edge. To bound the congestion, we hope that the probability in (33) can balance the number of possibilities. First note that P cannot be too large because the total number of paths $\mathcal{E}^*, (\mathcal{E}_i)_{i\in[s] \text{ is odd}}$ is at most $1 + \lceil \ell/2 \rceil$. Also note that if γ has P paths touching level $\ell - 1$, then σ also has at least P paths touching level $\ell - 1$. If we sample a random $\sigma \sim \mu^{ra}$, then the probability of having P such paths also decays fast with respect to P.

In the detailed analysis we need to deal with the summation over all (γ, γ') in (33). We first change the probability space such that instead of considering a random σ, we consider a random γ. Given a coloring γ, we reconstruct all $\mathcal{E}^*, (\mathcal{E}_i)_{i\in[s] \text{ is odd}}$ using γ so that we recover all possible initial colorings σ. We also need to find all possible next coloring γ' such (γ, γ') changes the color of a leaf edge e. The choices of γ' is also limited because there must exist a path in $\mathcal{E}^*, (\mathcal{E}_i)_{i\in[s] \text{ is odd}}$ that ends at e so that the canonical path will change the color of e.[1] Finally, we show that the γ with many possible σ and γ' appears with low probability and we can bound the congestion.

Detailed Analysis: Fix a leaf edge $h \in L_\ell$. Let $\gamma \in \Omega(\mu)$. We use $\mathcal{L}_h^\gamma$ to denote the set of available colors for edge h given coloring $\gamma(\mathbb{T}^* \setminus h)$. Let $c \in \mathcal{L}_h^\gamma \setminus \gamma(h)$. The coloring γ' can be written as $\gamma \oslash_h c$, which is the coloring obtained from γ by replacing the color on h with c. This is because the maximal $(\gamma(h), c)$-alternating path only has one edge h. Let $p^{ra} := 1/|\Omega(\mu^{ra})|$. In the distribution μ^{ra}, every coloring $x \in \Omega(\mu^{ra})$ has probability p^{ra} as μ^{ra} is the uniform distribution. Similarly, in

---

[1] By explicitly specifying the ordering used in Stage-I, one can even show that given γ, there is at most one possible γ' with $\gamma \oplus \gamma' \in L_\ell$ such that the move (γ, γ') can appear in the canonical path. However, we do not need to use this tight bound in our proof.

μ, each coloring $y \in \Omega(\mu)$ has probability $p := 1/ |\Omega(\mu)|$. We note that p^{ra} and p are two numbers rather than distributions. Note that $Q_\mu^{\text{Glauber}}(\gamma, \gamma \oslash_h c) = \frac{1}{q-d}$ where $d = \Delta - 1$. Since $\mathcal{C}_{ab}$ is the flip coupling, τ is fixed by σ and b. We simply denote $\gamma^{\sigma, \tau}$ by $\gamma^{\sigma, b}$ to emphasize that the path is fixed once σ and b are given. We have

$$
\begin{aligned}
\xi_\ell^{ab} &= \frac{q-d}{p} \sum_{\gamma \in \Omega(\mu)} \sum_{h \in L_\ell} \sum_{c \in \mathcal{L}_h^\gamma \backslash \gamma(h)} \left(\sum_{\sigma \in \Omega(\mu^{ra})} p^{ra} \cdot \mathbf{1}[(\gamma, \gamma \oslash_h c) \in \gamma^{\sigma, b}] \right)^2 \\
&= \frac{(q-d)(p^{ra})^2}{p^2} \sum_{\gamma \in \Omega(h)} p \sum_{h \in L_\ell} \sum_{c \in \mathcal{L}_h^\gamma \backslash \gamma(h)} \left| \left\{ \sigma \in \Omega(\mu^{ra}) \mid (\gamma, \gamma \oslash_h c) \in \gamma^{\sigma, b} \right\} \right|^2.
\end{aligned}
$$

Note that $\frac{p^{ra}}{p} = 1/\mu_r(a) = 2$ and $q = d + 3$. We have

$$
\frac{\xi_\ell^{ab}}{12} = \mathbb{E}_{\gamma \sim \mu} \left[\sum_{h \in L_\ell} \sum_{c \in \mathcal{L}_h^\gamma \backslash \gamma(h)} \left| \left\{ \sigma \in \Omega(\mu^{ra}) \mid (\gamma, \gamma \oslash_h c) \in \gamma^{\sigma, b} \right\} \right|^2 \right] := \mathcal{R}^{ab}. \tag{34}
$$

In the rest of this section, we construct canonical paths to make $\mathcal{R}^{ab} \le \frac{1}{24(\ell+1)}$ for all a, b, which implies $\xi_\ell \le \frac{1}{2(\ell+1)}$. Comparing to ξ_ℓ^{ab}, $\mathcal{R}^{ab}$ changes the randomness of expectation from the randomness of $(\sigma, \tau) \sim \mathcal{C}_{ab}$ to the randomness of $\gamma \sim \mu$, which will make the analysis easier. We analyze $\mathcal{R}^{ab}$ in the rest of this section.

Fix $\gamma \in \Omega(\mu)$. Without loss of generality, we assume the color of $\gamma(r)$ is either a or b. If $\gamma(r) = a$, then let $\mathcal{E}^*(\gamma)$ denote the longest (a, b)-alternating path starting from r. If $\gamma(r) = b$, then let $\mathcal{E}^*(\gamma)$ denote the longest b-a alternating path starting from r. We list all vertices along $\mathcal{E}^*(\gamma)$ as $u_0, u_1, \ldots, u_{S(\gamma)+1}$, where $r = \{u_0, u_1\}$ and $S(\gamma) + 1$ is the length of $\mathcal{E}^*$. In other words, $S(\gamma)$ is the number of edges in $\mathcal{E}^*(\gamma)$ except the root r. For every edge $f_i = \{u_i, u_{i+1}\} \in \mathcal{E}^*(\gamma)$, where $i \in [S(\gamma)]$ is odd, we define $\mathcal{E}_i(\gamma)$ be the path obtained by applying the same procedure for finding the path $\mathcal{E}_i$ in Stage-I. If $\gamma(r) = a$, it is exactly in the same setting as the procedure described in Stage-I. If $\gamma(r) = b$, one can go through the procedure in Stage-I to find the path.

For each $i \in [\ell]$ such that i is odd, we use $P_i(\gamma) \in \{0, 1\}$ to indicate the event that $i \le S(\gamma)$ and the $\mathcal{E}_i(\gamma) \cap L_{\ell-1}(\mathbb{T}^*) \ne \varnothing$. In words, $P_i(\gamma) = 1$ means the alternative path $\mathcal{E}^*(\gamma)$ has length at least i (so that e_i and $\mathcal{E}_i(\gamma)$ exist) and the path $\mathcal{E}_i(\gamma)$ contains some edge in level $\ell - 1$ of the tree $\mathbb{T}^*$. Define the sum $P(\gamma)$ as follows

$$
P(\gamma) := \sum_{i \in [\ell] : i \text{ is odd}} P_i(\gamma). \tag{35}
$$

Define $Z(\gamma)$ to indicate whether $\mathcal{E}^*(\gamma)$ touches level $\ell - 1$ in the tree $\mathbb{T}^*$, formally,

$$
Z(\gamma) := \mathbf{1}[S(\gamma) \ge \ell - 1]. \tag{36}
$$

Lemma 31. *Let $\Delta \ge 2$ and $q = \Delta + 2$ be two integers. Fix a and b. If $\ell \ge 1$ is odd, then for any $\gamma \in \Omega(\mu)$,*

$$
\sum_{h \in L_\ell} \sum_{c \in \mathcal{L}_h^\gamma \backslash \gamma(h)} \left| \left\{ \sigma \in \Omega(\mu^{ra}) \mid (\gamma, \gamma \oslash_h c) \in \gamma^{\sigma, b} \right\} \right|^2 \le 2(P(\gamma) + Z(\gamma)) \cdot (2\Delta)^{2(P(\gamma)+Z(\gamma))}.
$$

The proof of Lemma 31 is deferred to Section 5.3. We assume ℓ is odd in Lemma 31 because of the following two reasons. (1) To prove Lemma 19, we only need to show such ℓ exists. The existence of an odd ℓ is sufficient for us. (2) We can avoid discussing some corner cases when proving Lemma 31 if ℓ is odd. Hence, we use this assumption only for the simplicity of the proof. It is not difficult to generalize our proof in Section 5.3 to make it work for all ℓ.

Lemma 32. *Let $\Delta \geq 2$ and $q = \Delta + 2$ be two integers. Then, for any $0 \leq s \leq \ell$ and $0 \leq x \leq \lceil s/2 \rceil$,*

$$\mathbf{Pr}_{\gamma \sim \mu}[S(\gamma) = s \wedge P(\gamma) = x] \leq \left(1 - \frac{1}{\Delta}\right)^s \binom{\lceil s/2 \rceil}{x} \prod_{i=1}^{x} \left(1 - \frac{2}{\Delta}\right)^{\ell + 2i - s - 3}.$$

The proof of Lemma 32 is deferred to Section 5.4. Intuitively, there is an exponential decay of rate $1 - 1/\Delta$ which prevents the length of $\mathcal{E}^*(\gamma)$ and $P(\gamma)$ from being too large. Given Lemmas 31 and 32, we are ready to bound $\mathcal{R}^{ab}$ in (34). Now, we always assume ℓ is an odd number so that we can use Lemma 31. By Lemma 31, it holds that

$$\mathcal{R}^{ab} \leq \mathbb{E}_{\gamma \sim \mu}\left[2(P(\gamma) + Z(\gamma)) \cdot (2\Delta)^{2(P(\gamma) + Z(\gamma))}\right]$$

$$= \sum_{s=0}^{\ell} \sum_{x=0}^{\lceil s/2 \rceil} \mathbf{Pr}_{\gamma \sim \mu}[S(\gamma) = s \wedge P(\gamma) = x] \cdot 2(x + \mathbf{1}[s \geq \ell - 1])(2\Delta)^{2(x + \mathbf{1}[s \geq \ell - 1])} := \sum_{s=0}^{\ell} \sum_{x=0}^{\lceil s/2 \rceil} F(s, x).$$

The equation holds because it enumerates all possible values for $S(\gamma)$ and $P(\gamma)$. Note that $\binom{\lceil s/2 \rceil}{x} \leq \ell^x$. We can simplify the bound in Lemma 32 as

$$\mathbf{Pr}_{\rho \sim \mu}[S(\gamma) = s \wedge P(\gamma) = x] \leq \exp\left(-\frac{1}{\Delta}x^2 + \left(\ln \ell - \frac{\ell - s - 2}{\Delta}\right)x - \frac{s}{\Delta}\right). \tag{37}$$

Suppose $s \geq \ell - 1$. Using Lemma 32, we have

$$F(s, x) \leq 2(x + 1)(2\Delta)^{2(x+1)} \exp\left(-\frac{1}{\Delta}x^2 + \left(\ln \ell + \frac{2}{\Delta}\right)x - \frac{\ell - 1}{\Delta}\right)$$

$$\leq 2(\ell + 1) \exp\left(-\frac{1}{\Delta}x^2 + \left(\ln \ell + \frac{2}{\Delta} + 2\ln(2\Delta)\right)x + 2\ln(2\Delta) - \frac{\ell - 1}{\Delta}\right)$$

Note that $g(x) = -\frac{1}{\Delta}x^2 + \left(\ln \ell + \frac{2}{\Delta} + 2\ln(2\Delta)\right)x + 2\ln(2\Delta) = O(\Delta(\ln \ell + \ln \Delta)^2)$. We can take $\ell = O(\Delta^2 \log^2 \Delta)$ to make $F(s, x) \leq \frac{1}{100(\ell+1)^3}$.

Suppose $s < \ell - 1$, by rearranging terms in (37), we have

$$F(s, x) \leq 2x(2\Delta)^{2x} \exp\left(-\frac{1}{\Delta}x^2 + \left(\ln \ell - \frac{\ell - 2}{\Delta}\right)x + \frac{s}{\Delta}(x - 1)\right).$$

If $x \leq 1$, then $F(s, x) \leq F(0, 1) = 8\Delta^2 \exp(-\frac{\ell - 3}{\Delta} + \ln \ell) \leq \frac{1}{100(\ell+1)^3}$ for large $\ell = O(\Delta^2 \ln^2 \Delta)$. If $x > 1$, then $F(s, x) \leq F(\ell, x) = 2x(2\Delta)^{2x} \exp\left(-\frac{1}{\Delta}x^2 + \left(\ln \ell + \frac{2}{\Delta}\right)x - \frac{\ell}{\Delta}\right)$. Then, the upper bound $F(s, x) \leq \frac{1}{100(\ell+1)^3}$ can be obtained from a similar analysis for the case $s \geq \ell - 1$. Hence, there exists an odd $\ell = O(\Delta^2 \log^2 \Delta)$ such that

$$\mathcal{R}^{ab} \leq \sum_{s=0}^{\ell} \sum_{x=0}^{\lceil s/2 \rceil} F(s, x) \leq \sum_{s=0}^{\ell} \sum_{x=0}^{\lceil s/2 \rceil} \frac{1}{100(\ell + 1)^3} \leq \frac{1}{100(\ell + 1)}.$$

Finally, by (34), we have $\xi_\ell^{ab} \leq 12\mathcal{R}^{ab} < \frac{1}{2(\ell+1)}$ for all a, b.

5.3 Recovering the initial and final colorings: Proof of Lemma 31

For any initial σ and final τ, let $\gamma^{\sigma,\tau}$ be the canonical path. Since we assume ℓ is an odd number, the canonical path cannot change the color of a leaf edge at Stage-II. If (γ, γ') changes the color of the leaf edge, the move (γ, γ') is either in Stage-I or Stage-III. We can distinguish Stage-I or Stage-III by looking at the color of $\gamma(r) = \gamma'(r)$. If the color of the root hanging edge r is a, then the move is in Stage-I. Otherwise, the color of r is b and the move is in Stage-III. We first assume (γ, γ') is in Stage-I. We show how to recover the initial state σ. Note that since a and b are fixed. Once we recover σ, the τ is uniquely fixed. The other case is (γ, γ') is in Stage-III. We will discuss it at the end of the proof.

Let $\gamma \in \Omega(\mu)$ with $\gamma(r) = a$. Let $h \in L_\ell(\mathbb{T}^\star)$ be a leaf node. Let c be an available color for h in γ. Given a move $(\gamma, \gamma \oslash_h c)$, we want to recover how many possible pairs σ such that the canonical path $\gamma^{\sigma,\tau}$ uses the move, where $\tau = \sigma \oslash_r b$ is fixed. Recall $\mathcal{E}^*(\gamma)$ is the longest (a,b)-alternating path in γ, with vertex sequence $u_0, u_1, \ldots, u_{S(\gamma)+1}$. Recall $\mathcal{E}_i(\gamma)$ is defined in Section 5.2.

Lemma 33. *Fix a coloring $\gamma \in \Omega(\mu)$ with $\sigma(r) = a$. For any $h \in L_\ell(\mathbb{T}^\star)$ and any $c \in \mathcal{L}_h^\gamma \setminus \gamma(h)$, the move $(\gamma, \gamma \oslash_h c)$ can appear in some canonical path $\gamma^{\sigma,\tau}$ only if one of the following two conditions holds*

- $h \in \mathcal{E}^*(\gamma)$;

- *there exists an odd $i \in [S(\gamma)]$ with $i \neq \ell$ such that $h \in \mathcal{E}_i(\gamma)$.*

Proof. Suppose the move $(\gamma, \gamma \oslash_h c)$ appears in some canonical path $\Gamma^{\sigma,\tau}$. It must appear in Stage-I. In Stage-I, the algorithm uses a specific ordering to change the colors of edges. It holds that $\gamma(h) = \sigma(h)$ and for all non-leaf edge e, $\gamma(e) = \sigma(e)$. Let $\mathcal{E}^*(\sigma) = \sigma \oplus \tau$ be the (a,b)-alternating path for the initial coloring σ. Recall $\mathcal{E}^*(\sigma)$ has vertices $v_0, v_1, \ldots, v_{S(\sigma)+1}$. Let $\mathcal{E}_i(\sigma)$ denote the path $\mathcal{E}_i$ constructed in Stage-I. Since the move is used in the canonical path, there are only two cases.

- The leaf edge h is the last edge in $\mathcal{E}^*(\sigma)$. In this case $\mathcal{E}^*(\sigma) = \mathcal{E}^*(\gamma)$ and $h \in \mathcal{E}^*(\gamma)$.

- The leaf edge h belongs to $\mathcal{E}_j(\sigma)$, where $j \in [S(\sigma)]$ is odd and $j \leq \ell - 2$. In this case $\mathcal{E}_j(\gamma) = \mathcal{E}_j(\sigma)$, because no edges in $\mathcal{E}_j(\sigma)$ changed its color in the path from σ to γ. Hence, $h \in \mathcal{E}_j(\gamma)$.

In the second case, we assume $j \leq \ell - 2$. This is because both j and ℓ are odd numbers. Also note that $j \neq \ell$, because any leaf edge must have an available color $c \notin \{a, b\}$ and $\mathcal{E}_\ell(\sigma) = \varnothing$. $\square$

Fix a coloring $\gamma \in \Omega(\mu)$ with $\sigma(r) = a$. Fix $h \in L_\ell(\mathbb{T}^\star)$ and $c \in \mathcal{L}_h^\gamma \setminus \gamma(h)$ satisfying the condition in Lemma 33. For any odd $i \in [S(\gamma)]$ with $i \leq \ell - 2$, we say the path $\mathcal{E}_i(\gamma)$ is *suspicious* if the last edge is in the level $\ell - 1$ of $\mathbb{T}$, formally, $\mathcal{E}_i(\gamma) \cap L_{\ell-1}(\mathbb{T}^\star) \neq \varnothing$ and $\mathcal{E}_i(\gamma) \cap L_\ell(\mathbb{T}^\star) = \varnothing$. We use $i_1, i_2, \ldots, i_\beta$ to list the indices of all suspicious paths. For any $j \in [\beta]$, we use $H(i_j)$ to denote the set of leaf edges adjacent to the last edge in path $\mathcal{E}_{i_j}(\gamma)$. Finally, we say the alternating path $\mathcal{E}^*(\gamma)$ is *suspicious* if $S(\gamma) = \ell - 1$. If $\mathcal{E}^*(\gamma)$ is suspicious, let H_0 be the set of leaf edges adjacent to $\mathcal{E}^*(\gamma)$; otherwise, let $H_0 = \varnothing$. By the definition of paths $\mathcal{E}_i(\sigma)$ and $\mathcal{E}^*(\sigma)$, it is easy to verify that

- $H_0, H(i_1), \ldots, H(i_\beta)$ are disjoint.

- Since the degree of the tree is Δ, we have $|H_0| < \Delta$ and $|H(i_j)| < \Delta$ for all $j \in [\beta]$.

Next, we show the following results.

Lemma 34. *The move $(\gamma, \gamma \oslash_h c)$ appears in a canonical path $\gamma^{\sigma,\tau}$ only if*

- $\sigma \oplus \gamma \subseteq H_0 \cup H(i_1) \cup \ldots \cup H(i_\beta)$;

- *For any $H \in \{H_0, H(i_1), \ldots, H(i_\beta)\}$, $|(\sigma \oplus \gamma) \cap H| \leq 1$.*

Proof. Consider a canonical path $\Gamma^{\sigma, \tau}$ where σ and τ come from the flipping coupling. Since we assume $\gamma(r) = a$, the move $(\gamma, \gamma \oslash_h c)$ can only occur at the Stage-I. Consider all the paths $\mathcal{E}_i(\sigma)$, where $i \in [S(\sigma)]$ is odd and $i \neq \ell$. If $\mathcal{E}_i(\sigma)$ contains a leaf edge, then we denote this leaf edge by $h(i)$. If $S(\sigma) = \ell$, then we use $h(\ell)$ to denote the last leaf edge in $\mathcal{E}^*(\sigma)$. Let $j_1, j_2, \ldots, j_\alpha$ be all odd indices in $[S(\sigma)]$ such that the leaf edge $h(j_k)$ is defined. Then, all these α leaf edges are the first α edges in $\mathcal{E}$ according to the ordering in Stage-I. Assume the first α steps in Stage-I are to change the colors of edges $h(j_1), h(j_2), \ldots, h(j_\alpha)$ so that we obtain a path $(\sigma = \gamma_0, \gamma_1, \ldots, \gamma_\alpha)$.

Fix $1 \leq k \leq \alpha$. Recall $S(\gamma_k)$ denotes the length of the maximal (a, b)-alternating path $\mathcal{E}^*(\gamma_k)$ from r in γ_k. For any odd $i \in [S(\gamma_k)]$, again we can define the path $\mathcal{E}_i(\gamma_k)$ (applying Stage-I from γ_k to construct paths $\mathcal{E}_i(\gamma_k)$). We claim the following properties.

- for any $1 \leq k' \leq k-1$ with $j_{k'} \neq \ell$, it holds that $\mathcal{E}_{j_{k'}}(\sigma) = (\mathcal{E}_{j_{k'}}(\gamma_k), h(j_{k'}))$. In words, $\mathcal{E}_{j_{k'}}(\gamma_k)$ is a suspicious path obtained from $\mathcal{E}_{j_{k'}}(\sigma)$ by removing the last edge (the leaf edge).

- if $j_{k^*} = \ell$ for some $1 \leq k^* \leq k-1$, then $S(\gamma_k) = \ell - 1$.

The first property holds because once we change the color of $h(j_{k'})$, $\mathcal{E}_{j_k}(\sigma)$ becomes the suspicious path $\mathcal{E}_{j'_k}(\sigma) = \mathcal{E}_{j_k}(\sigma) \setminus h(j_{k'})$ and $\mathcal{E}_{j_k}(\sigma) = \mathcal{E}_{j'_k}(\sigma)$. The second property holds because once we change the color of the leaf edge in the alternating path $\mathcal{E}^*(\sigma)$, then $\mathcal{E}^*(\gamma_{j_{k^*}}) = \mathcal{E}^*(\sigma) \setminus h(j_{k^*})$, which means the length of the alternating path decreases by 1 and we have $S(\gamma_k) = S(\gamma_{k^*}) = \ell - 1$. The above two properties says the first $k-1$ moves are "recorded" in γ_k by suspicious paths and $S(\gamma_k)$. Formally, let R be the set of all indices of suspicious paths in γ_k. It holds that $\{j_1, j_2, \ldots, j_{k-1}\} \subseteq R \cup \{\ell\}$ and $\ell \in \{j_1, j_2, \ldots, j_{k-1}\}$ only if $S(\gamma_k) = \ell - 1$.

Now given a move $(\gamma, \gamma \oslash_h c)$, we show how to recover all possible starting colorings σ. The σ must have the same colors as γ for all edges at level $\leq \ell - 1$. We list all the indices $i_1, i_2, \ldots, i_\beta$ of suspicious paths for γ. Let $H(i_k)$ and H_0 be the sets in the lemma. By above analysis, for any $k \in [\beta]$, in σ, either $\sigma(H(i_k)) = \gamma(H(i_k))$ (the case $\mathcal{E}_{i_k}(\sigma) = \mathcal{E}_{i_k}(\gamma)$) or there exists only one edge $h(i_k) \in H(i_k)$ such that $\sigma(H(i_k))$ differs from $\gamma(H(i_k))$ at the color of $h(i_k)$ (the case $\mathcal{E}_{i_k}(\sigma) = (\mathcal{E}_{i_k}(\gamma), h(i_k))$). If $S(\gamma) = \ell - 1$, then σ and γ can differ at ≤ 1 edge in H_0. For all other leaf edges not in $H_0 \cup H(i_1) \cup \ldots \cup H(i_\beta)$, σ and γ must take the same color. This proves the lemma. $\qquad\square$

Now, we are ready to prove Lemma 31. Fix a γ with $\gamma(r) = a$. By Lemma 33, the number of possible $h \in L_\ell(\mathbb{T}^*)$ are at most $P(\gamma) + Z(\gamma)$. For each h, the number of $c \in \mathcal{L}_h^\gamma \setminus \gamma(h)$ is at most $|\mathcal{L}_h^\gamma \setminus \gamma(h)| = q - \Delta = 2$.[2] Next, given $(\gamma, \gamma \oslash_h c)$, we need to recover the number of possible σ's. By Lemma 34, for each $H \in \{H_0, H(i_1), \ldots, H(i_\beta)\}$, either $\sigma(H) = \gamma(H)$ or there exists $h \in H$, such that γ and σ differ only at h, and to get a proper coloring, the number of possible colors for $\sigma(h)$ is $q - \Delta = 2$. Recall $|H| \leq \Delta - 1$ for all $H \in \{H_0, H(i_1), \ldots, H(i_\beta)\}$. Hence, the total number of possibilities is $1 + 2(\Delta - 1) < 2\Delta$. Note that $H_0 \neq \varnothing$ only if $Z(\gamma) \geq 1$.

$$\sum_{h \in L_\ell} \sum_{c \in \mathcal{L}_h^\gamma \setminus \gamma(h)} \left| \left\{ \sigma \in \Omega(\mu^{ra}) \mid (\gamma, \gamma \oslash_h c) \in \gamma^{\sigma, b} \right\} \right|^2 \leq 2(P(\gamma) + Z(\gamma)) \cdot (2\Delta)^{2(X(\gamma) + Z(\gamma))}.$$

Finally, consider the case $\gamma(r) = b$, and thus (γ, γ') is in Stage-III. By Claim 30, Stage-III is the reverse of the Stage-I starting from τ. Hence, Stage-III changes the colors of the leaves at the end of the process. Instead of recovering the initial coloring σ, we recover the final coloring τ in this case. The proof in this case follows by symmetry.

---

[2] We remark that one can further reduce the number of $c's$ to 1 because the algorithm always uses the first available color in the ordering $\mathcal{O}$. But the upper bound 2 is enough for us.

5.4 Bounding the probability of reaching the leaves: Proof of Lemma 32

Now, we prove Lemma 32. Suppose $S(\gamma) = s$ and $P(\gamma) = x$. We can fix the x odd indices in $[s]$ such that $P_i(\gamma) = 1$. There all at most $\binom{\lceil s/2 \rceil}{x}$ ways to choose these indices. Fix a choice $j_1 > j_2 > \ldots > j_x$. We bound the probability of $S(\gamma) = s$ and $P_{j_i}(\gamma) = 1$ for all $i \in [x]$. Consider the following process, we first sample a random $\gamma \in \mu$ and then reveal the colors on edges one by one. We first reveal $\gamma(r)$. Without loss of generality, we assume $\gamma(r) = a$. The case $\gamma(r) = b$ follows by symmetry. We then reveal colors of all children edges of r. If one child edge e takes the color b, which happens with probability $\frac{\Delta - 1}{q - 1}$, we reveal the colors of all children edges of e and look for the color a. If $S(\gamma) = s$, we must repeat this process for s steps and reveal colors on a subset Λ. It gives

$$\mathbf{Pr}\left[S(\gamma) = s\right] \le \left(\frac{\Delta - 1}{q - 1}\right)^s \le \left(1 - \frac{q - \Delta}{q - 1}\right)^s \le \left(1 - \frac{1}{\Delta}\right)^s.$$

For any $i \in [x]$, we simulate the process in Stage-I for finding the path $\mathcal{E}_i$. Given the coloring in Λ, we either know $\mathcal{E}_{j_i} = \varnothing$ or we know the first edge $(u_1 = v_{j_i}, u_2)$ in $\mathcal{E}_{j_i}$. This is because the coloring on Λ gives the colors of all edges adjacent to the (a,b)-alternating path. If $\mathcal{E}_{j_i} = \varnothing$, then the probability of $P_{j_i}(\gamma) = 1$ is 0. Assume the worst case that the first edge (u_1, u_2) is fixed. At this moment, we already revealed the colors of all edges incident to u_1. Since $q = \Delta + 2$, there are 2 missing colors, say colors c and c'. We will keep constructing the path $\mathcal{E}_{j_i}$ only if (u_1, u_2) has less than 2 available colors, which happens only if both c and c' appear in child-edges of u_2. We reveal in coloring of child-edges of u_2, conditional on all information revealed so far,[3] this happens with probability $\binom{q-3}{\Delta-1}/\binom{q-1}{\Delta-1} = \frac{\Delta-2}{\Delta}$. This event needs should happen for $\ell - 1 - j_1$ times if $P_{j_i}(\gamma) = 1$. The probability is at most $(1 - \frac{2}{\Delta})^{\ell - 1 - j_1}$. Finally, we can use this argument for all paths $\mathcal{E}_{j_i}$ for $i \in [x]$, and all the edges we considered above are disjoint. This gives

$$\mathbf{Pr}\left[S(\gamma = s) \wedge \forall i \in [x], P_{j_i}(\gamma) = 1\right] \le \left(1 - \frac{1}{\Delta}\right)^s \prod_{i=1}^{x}\left(1 - \frac{2}{\Delta}\right)^{\ell - 1 - j_i}$$

$$(\text{by } j_i \ge s - 2(i-1)) \quad \le \left(1 - \frac{1}{\Delta}\right)^s \prod_{i=1}^{x}\left(1 - \frac{2}{\Delta}\right)^{\ell + 2i - s - 3}.$$

The above bound relies on the fact the graph is a tree, so that we can use conditional independence property to take the product of all probabilities. The lemma follows from a union bound over all possible $j_1, j_2, \ldots, j_x$.

6 Coupling and canonical paths for the edge dynamics

In this section, we give a proof overview for Lemma 23. The proof will follow the same high-level plan described in Section 3.2 and Section 5. We assume the readers are already familiar with those sections. For convenience, we denote $\mathbb{T}_\ell^\star, \mu_\ell^\star$ as $\mathbb{T}$ and μ, respectively. Also, we will use $\widetilde{\mathbb{T}}$ to denote the tree $\mathbb{T} - r$.

Recall that the collection of available update is defined in $\mathcal{B}$ in (17). For convenient, we will use

$$\mathcal{A} := \{\{e, r\} \mid e \in L_1(\mathbb{T}_\ell^\star)\} \tag{38}$$

$$\mathcal{B} = \{\{e\} \mid e \in \mathbb{T}_\ell^\star\} \cup \mathcal{A}. \tag{39}$$

---

[3]Let Λ denote the set of edges revealed so far. The Λ is a random subset. However, the random Λ is fully determined by $\gamma(\Lambda)$. Hence, given Λ and $\gamma(\Lambda)$, all other variables in $\overline{\Lambda}$ follows the distribution of μ conditional on $\gamma(\Lambda)$.

Our goal is to establish approximate root-factorization of variance as defined in (18). In particular, we need to verify it with parameter α and β as in Lemma 23. Recall this states that for any function $f : \Omega(\mu) \to \mathbb{R}$, the following holds:

$$\mathbf{Var}_\mu \left[\mu_{\widehat{\mathbb{T}}}[f] \right] \leq \left(\sum_{i=0}^{\ell} \alpha_i \sum_{h \in L_i} \mu[\mathbf{Var}_h [f]] \right) + \beta \sum_{B \in \mathcal{A}} \mu[\mathbf{Var}_B [f]]. \tag{40}$$

To allow the update on multiple edges, we generalize some definitions in Section 3.2.

Definition 35 (Canonical paths w.r.t. a coupling set). Given a coupling set $\mathcal{C}$, for any pair $(\sigma, \tau) \in \Omega(\mathcal{C})$, a canonical path $\gamma^{\sigma,\tau}$ is a simple path from σ to τ on the graph $(\Omega(\mu), \Lambda(\mu))$, where an edge $(\gamma, \gamma') \in \Lambda(\mu)$ if and only if $\gamma \oplus \gamma' \in \mathcal{B}$.
   Let $\Gamma = \{\gamma^{\sigma,\tau} \mid (\sigma, \tau) \in \Omega(\mathcal{C})\}$ denote the set of canonical paths w.r.t. the coupling set $\mathcal{C}$.

In the following, we use $Q_\mu^{\mathcal{B}}(\gamma, \gamma')$ to denote the transition rate from γ to γ' (with $\gamma \oplus \gamma' = B \in \mathcal{B}$) in the (continuous-time) heat-bath $\mathcal{B}$-block dynamics for μ. Formally,

$$Q_\mu^{\mathcal{B}}(\gamma, \gamma') = \mu^{\gamma_{\mathbb{T} \setminus B}}(\gamma') = \mu_B^{\gamma_{\mathbb{T} \setminus B}}(\gamma'_B).$$

Definition 36 (Congestion of canonical paths w.r.t. a coupling set). Given a coupling set $\mathcal{C}$ and a set of canonical paths Γ, define the expected congestion at level $0 \leq t \leq \ell$ by

$$\xi_t = \max_{\substack{a,b \in [q-d]: \\ a \neq b}} \sum_{\substack{s=(\gamma,\gamma') \in \Lambda(\mu): \\ (\gamma \oplus \gamma') \in L_t}} \frac{\left(\mathbf{Pr}_{(\sigma,\tau) \sim \mathcal{C}_{ab}} [s \in \gamma^{\sigma,\tau}] \right)^2}{\mu(\gamma) Q_\mu^{\mathcal{B}}(\gamma, \gamma')}. \tag{41}$$

Then, define the expected congestion at $\mathcal{A}$ by

$$\xi_{\mathcal{A}} = \max_{\substack{a,b \in [q-d]: \\ a \neq b}} \sum_{\substack{s=(\gamma,\gamma') \in \Lambda(\mu): \\ (\gamma \oplus \gamma') \in \mathcal{A}}} \frac{\left(\mathbf{Pr}_{(\sigma,\tau) \sim \mathcal{C}_{ab}} [s \in \gamma^{\sigma,\tau}] \right)^2}{\mu(\gamma) Q_\mu^{\mathcal{B}}(\gamma, \gamma')}. \tag{42}$$

Note that in (41) the summation is over pairs γ, γ' that only differ at a single edge. In this case, it holds that $Q_\mu^{\mathcal{B}}(\gamma, \gamma') = Q_\mu^{\text{Glauber}}(\gamma, \gamma')$ and hence (41) has the same form as (11).
   Similar to Lemma 18, we have the following result.

Lemma 37. *If there exist a set of couplings $\mathcal{C}$ and a set Γ of canonical paths such that the congestion with respect to coupling is ξ_t for $0 \leq t \leq \ell$ and $\xi_{\mathcal{A}}$ for $\mathcal{A}$, then (40) holds with $\alpha = 2(\ell + 1)\xi$ and $\beta = 2\xi_{\mathcal{A}}$.*

Now, following the proof of Lemma 18, we have,

Proof. We note that by definition

$$\mathbf{Var}_\mu \left[\mu_{\widehat{\mathbb{T}}}[f] \right] = \frac{1}{2} \sum_{a,b \in [q-d]: a \neq b} \mu_r(a)\mu_r(b) \left(\mathbb{E}_{(\sigma,\tau) \sim \mathcal{C}_{ab}} \left[\sum_{j=1}^{m(\sigma,\tau)} f(\gamma_j^{\sigma,\tau}) - f(\gamma_{j-1}^{\sigma,\tau}) \right] \right)^2.$$

Fix a and b. Denote $\mathbb{E}_{(\sigma,\tau) \sim \mathcal{C}_{ab}} \left[\sum_{j=1}^{m(\sigma,\tau)} f(\gamma_j^{\sigma,\tau}) - f(\gamma_{j-1}^{\sigma,\tau}) \right]$ by $\widehat{E}_{ab}$. Then $\widehat{E}_{ab}$ can be rewritten as

$$\widehat{E}_{ab} = E_{ab} + \sum_{\substack{(\gamma,\gamma'): \\ \gamma \oplus \gamma' \in \mathcal{A}}} (f(\gamma) - f(\gamma')) \mathbf{Pr}_{(\sigma,\tau) \sim \mathcal{C}_{ab}} \left[\gamma^{\sigma,\tau} \text{ uses } (\gamma, \gamma') \right],$$

30

where E_{ab} is defined in (12). By using Cauchy-Schwarz,

$$\widehat{E}_{ab}^2 \le 2E_{ab}^2 + 2\left(\sum_{\substack{(\gamma,\gamma'): \\ \gamma \oplus \gamma' \in \mathcal{A}}} (f(\gamma) - f(\gamma'))\mathbf{Pr}_{(\sigma,\tau)\sim\mathcal{C}_{ab}} \left[\gamma^{\sigma,\tau} \text{ uses } (\gamma,\gamma') \right] \right)^2 := 2E_{ab}^2 + 2A^2.$$

We note that E_{ab} is already bounded in the proof of Lemma 18, we only need to bound A^2. Note that A and A_i in (13) has the same form. By the argument in the proof of Lemma 18, we have

$$A^2 \le 2\xi_{\mathcal{A}} \sum_{B\in\mathcal{A}} \mu[\mathbf{Var}_B [f]].$$

Combining everything together, we have

$$\mathbf{Var}_\mu \left[\mu_{\widetilde{\mathbb{T}}}[f] \right] \le 2 \sum_{a,b\in[q-d]:a\neq b} \mu_r(a)\mu_r(b) \left((\ell+1) \sum_{i=0}^{\ell} \xi_i \sum_{e\in L_i} \mu[\mathbf{Var}_e [f]] + \xi_{\mathcal{A}} \sum_{B\in\mathcal{A}} \mu[\mathbf{Var}_B [f]] \right)$$

$$\le 2(\ell+1) \sum_{i=0}^{\ell} \xi_i \sum_{e\in L_i} \mu[\mathbf{Var}_e [f]] + 2\xi_{\mathcal{A}} \sum_{B\in\mathcal{A}} \mu[\mathbf{Var}_B [f]].$$

Hence, the root-factorization of variance in (40) holds with $\boldsymbol{\alpha} = 2(\ell+1)\boldsymbol{\xi}$ and $\beta = 2\xi_{\mathcal{A}}$. □

6.1 Canonical paths and congestion analysis

With Lemma 37 in hand, we could bound α_i for $i < \ell$ and β by using the crude $q^{\Delta^{O(\ell)}}$ bound. In order to prove Lemma 23, we only need to verify $\alpha_\ell \le 1/2$ for some large ℓ. Just like what we did in Section 5. Recall that in the setting of Lemma 23, $q = \Delta + 1 = d + 2$. The major difference of this section from Section 5 is that we will use a slightly different construction of canonical path.

Fix $a, b \in [q-d]$ with $a \neq b$, we will use the same coupling $\mathcal{C}_{a,b}$ between μ^{ra} and μ^{rb} as we did in Section 5. Following the notations in Section 5, we can find the maximal (a,b)-alternating path $\mathcal{E}^* = (e_0, e_1, e_2, \cdots, e_s)$ of length s in σ such that $e_0 = r$; $\sigma(e_i) = a$ for even i and $\sigma(e_j) = b$ for odd j. The destination configuration τ is then obtained by exchanging the colors a and b in $\mathcal{E}^*$.

Note that since $q = \Delta + 1$, every edge e has at most 2 available colors when we fix the colors of its neighbors. This means when we say we "flip" the color of some edge, there is only one choice of its new color. Hence, in this section, we could omit the ordering $\mathcal{O}$ that is used in Section 5 for simplicity. Now, we start our construction of the canonical path $\gamma^{\sigma,\tau}$. The construction takes care of the parity of $|\mathcal{E}^*|$. We also assume ℓ is odd as in Section 5 to avoid annoying corner cases.

When $|\mathcal{E}^*|$ is odd or $|\mathcal{E}^*| = \ell+1$: In this case, we use the same construction for $\gamma^{\sigma,\tau}$ as in Section 5. The condition on $|\mathcal{E}^*| =: s$ ensures that for each odd $i \le \ell$, we have either

1. $i \neq s$, i.e., e_i is not the last edge in $\mathcal{E}^*$ (happens when $|\mathcal{E}^*|$ is odd);

2. $i = s$ but $e_s = e_i$ is a leaf of $\mathbb{T}$ (happens when $|\mathcal{E}^*| = \ell+1$ is even).

Now, for an odd i, in **Stage-I** of Section 5, the edge list $\mathcal{E}_i$ is constructed as follow.

When e_i has 2 available colors, we know that it must have an available color $c \notin \{a,b\}$, and hence we could let $\mathcal{E}_i = \varnothing$. We note that the case $i = s$ and e_i is a leaf falls into this case.

When e_i only has 1 available color, i.e., its current color, then e_i cannot be a leaf. By our assumption, this means e_i cannot be the last edge in $\mathcal{E}^*$. Suppose $e_i = (v_i, v_{i+1})$ such that v_{i+1} is a child of

31

v_i. Then the unique edge $e \in N_e(v_i) \setminus \{e_i\}$ such that $\sigma(e) \in [q] \setminus \sigma(N_e(v_{i+1}))$ will be add to $\mathcal{E}_i$. This process will be applied recursively on e until it finally reaches an edge with 2 available colors (e.g., leaves). We refer the reader to Section 5 for a more detailed description.

Here, a crucial observation is that $e \neq e_{i-1}$. This is because we know $\sigma(e_{i+1}) = a$ and hence $\sigma(e_{i-1}) = a \notin [q] \setminus \sigma(N_e(v_{i+1}))$. But we know that $\sigma(e) \in [q] \setminus \sigma(N_e(v_{i+1}))$ which means e cannot be e_{i-1}. This property ensures that for different odd i, j, the distance between $(e_i \cup \mathcal{E}_i)$ and $(e_j \cup \mathcal{E}_i)$ is at least 2 in the line graph. This means all the argument in Section 5 also works in this case.

Otherwise: In this case, the last edge e_s in $\mathcal{E}^*$ must be an odd edge, i.e., s is odd. If we use the same construction of $\gamma^{\sigma,\tau}$ as in Section 5, we can no longer ensure that for different odd i, j, the distance between $(e_i \cup \mathcal{E}_i)$ and $(e_j \cup \mathcal{E}_i)$ is at least 2. Hence the argument in Section 5 will fail. See Figure 3 for an example.
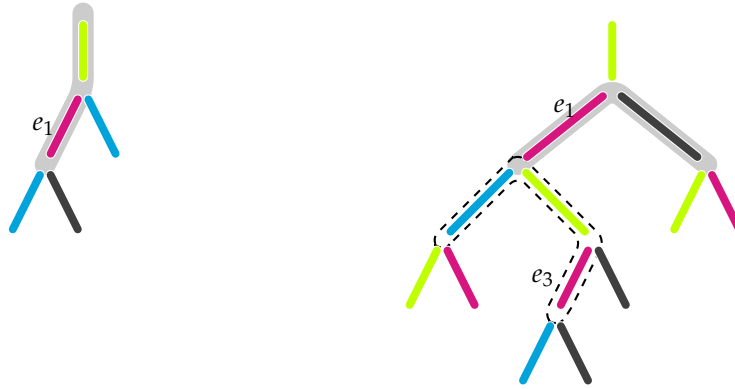

Figure 3: We consider two 3-regular trees with $\bullet, \bullet, \bullet, \bullet$ as avaliable colors. In both trees, we set $a = \bullet$ and $b = \bullet$. In the first example, we mark the odd numbered edge e_1 with corresponding $\mathcal{E}_1$ using shadowed background. This means, in order to change the color of e_1, by the construction of $\mathcal{E}_1$ we used in Section 5, we want to change the color of the root edge. Since the root edge only have two avaliable color $\bullet$ and $\bullet$, this is not possible.
In the second example, we mark e_1 and $\mathcal{E}_1$ with shadowed background and we mark e_3 and $\mathcal{E}_3$ with dashed circle. It is then easy to see that these paths have neiboring edges. This means, after fliping the edges in $e_3 \cup \mathcal{E}_3$, the path $e_1 \cup \mathcal{E}_1$ will be different. This will fail the construction in Section 5, which havily relies on these paths can be handled independently.

However, in this case, we note that all the even edge e_j cannot be the last edge of $\mathcal{E}^*$. Hence, we could modify the original construction of $\gamma^{\sigma,\tau}$ as follow.

Stage-I For every positive even edge j (excluding $e_0 = r$), construct $\mathcal{E}_j$ as we did for odd edges. Then we flip these edges ($\cup_{\text{pos. even } j}(\mathcal{E}_j \cup \{e_j\})$) from button up while making sure that edges in $\mathcal{E}_j$ is flipped before e_j. After that, each edge e_j have changed its color from a to $c_i \notin \{a, b\}$.

Stage-II Exchange the color of e_0 and e_1. Since $|\mathcal{E}^*|$ is even, we know that $|\mathcal{E}^*| \geq 2$. Then, change the color of all the other odd edge e_i from b to a.

Stage-III Reverse of **Stage-I**.

We note that **Stage-II** is the only place where we use the update $\{e_0 = r, e_1\}$. Recall that the definition of ξ_ℓ that we used in this section has exactly the same form as in Section 5. With this construction of $\gamma^{\sigma,\tau}$, one could then verifies $\alpha_\ell \leq 1/2$ and finish the proof of Lemma 23 by routinely

following the argument used in the proof of Lemma 31 and Lemma 32. We omit this part for simplicity.

7 Open problems

We provide some interesting directions to explore in future works.

1. For $q = \Delta + 1, \Delta \geq 3$, what is the relaxation time of the Glauber dynamics on Δ-regular trees? We have a lower bound of $\Omega(\Delta n)$ and an upper bound of $\Delta n^{1+O(1/\log \Delta)}$.

2. Same question for general trees of maximum degree Δ. For $\Delta = 2$ there is a mixing time lower bound of $\Omega(n^3 \log n)$ for $\Delta = 2$ by [DGJ06]. Is there a similar lower bound for general trees when $\Delta \geq 3$?

3. For $q \geq \Delta + 2$ can we get tight bounds on the mixing time for general trees?

4. For general graphs (not necessarily trees), [WZZ24] proved rapid mixing of the Glauber dynamics when $q > (2 + o(1))\Delta$. Is there a MCMC algorithm (or other method) when $q \geq \Delta + 2$, or is there a hardness of approximate counting result in this parameter region?

References

[AL20]   Vedat Levi Alev and Lap Chi Lau. Improved analysis of higher order random walks and applications. In *Proceedings of the 52nd Annual ACM Symposium on Theory of Computing (STOC)*, pages 1198–1211, 2020.

[ALG21]   Dorna Abdolazimi, Kuikui Liu, and Shayan Oveis Gharan. A matrix trickle-down theorem on simplicial complexes and applications to sampling colorings. In *Proceedings of the 62nd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 161–172, 2021.

[ALO20]   Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. Spectral independence in high-dimensional expanders and applications to the hardcore model. In *Proceedings of the 61st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 1319–1330, 2020.

[BCC$^+$22] Antonio Blanca, Pietro Caputo, Zongchen Chen, Daniel Parisi, Daniel Štefankovič, and Eric Vigoda. On mixing of Markov chains: Coupling, spectral independence, and entropy factorization. In *Proceedings of the 33rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 3670–3692, 2022.

[BCP$^+$22] Antonio Blanca, Pietro Caputo, Daniel Parisi, Alistair Sinclair, and Eric Vigoda. Entropy decay in the Swendsen-Wang dynamics on $\mathbb{Z}^d$. *Ann. Appl. Probab.*, 32(2):1018–1057, 2022.

[BKMP05] Noam Berger, Claire Kenyon, Elchanan Mossel, and Yuval Peres. Glauber dynamics on trees and hyperbolic graphs. *Probab. Theory Related Fields*, 131(3):311–340, 2005.

[Cap23]   Pietro Caputo. Lecture notes on Entropy and Markov Chains. *Preprint, available from: http://www.mat.uniroma3.it/users/caputo/entropy.pdf*, 2023.

[CDM$^+$19] Sitan Chen, Michelle Delcourt, Ankur Moitra, Guillem Perarnau, and Luke Postle. Improved bounds for randomly sampling colorings via linear programming. In *Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2216–2234, 2019.

[Che24]   Zongchen Chen. Combinatorial approach for factorization of variance and entropy in spin systems. In *Proceedings of the 35th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 4988–5012, 2024.

[CLV21]   Zongchen Chen, Kuikui Liu, and Eric Vigoda. Optimal mixing of Glauber dynamics: Entropy factorization via high-dimensional expansion. In *Proceedings of the 53rd Annual ACM Symposium on Theory of Computing (STOC)*, pages 1537–1550, 2021.

[CMT15]   Pietro Caputo, Georg Menz, and Prasad Tetali. Approximate tensorization of entropy at high temperature. *Ann. Fac. Sci. Toulouse Math. (6)*, 24(4):691–716, 2015.

[CP21]   Pietro Caputo and Daniel Parisi. Block factorization of the relative entropy via spatial mixing. *Communications in Mathematical Physics*, 388:793–818, 2021.

[CYYZ24]   Xiaoyu Chen, Xiongxin Yang, Yitong Yin, and Xinyuan Zhang. Spectral independence beyond total influence on trees and related graphs. *ArXiv preprint arXiv:2404.04668*, 2024.

[DFFV06]   Martin Dyer, Abraham D. Flaxman, Alan M. Frieze, and Eric Vigoda. Randomly coloring sparse random graphs with fewer colors than the maximum degree. *Random Structures & Algorithms*, 29(4):450–465, 2006.

[DGJ06]   Martin Dyer, Leslie Ann Goldberg, and Mark Jerrum. Systematic scan for sampling colorings. *The Annals of Applied Probability*, 16(1):185–230, 2006.

[DGU14]   Martin Dyer, Catherine Greenhill, and Mario Ullrich. Structure and eigenvalues of heat-bath Markov chains. *Linear Algebra and its Applications*, 454:57–71, 2014.

[DHP20]   Michelle Delcourt, Marc Heinrich, and Guillem Perarnau. The Glauber dynamics for edge-colourings of trees. *Random Structures & Algorithms*, 57(4):1050–1076, 2020.

[DS91]   Persi Diaconis and Daniel Stroock. Geometric bounds for eigenvalues of Markov chains. *The Annals of Applied Probability*, 1(1):36–61, 1991.

[DS96]   P. Diaconis and L. Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *The Annals of Applied Probability*, 6(3):695–750, 1996.

[EF23]   David Eppstein and Daniel Frishberg. Rapid mixing for the hardcore Glauber dynamics and other Markov chains in bounded-treewidth graphs. In *Proceedings of the 34th International Symposium on Algorithms and Computation (ISAAC)*, pages 30:1–30:13, 2023.

[EHSV23]   Charilaos Efthymiou, Thomas P. Hayes, Daniel Stefankovic, and Eric Vigoda. Optimal mixing via tensorization for random independent sets on arbitrary trees. In *APPROX/RANDOM*, pages 33:1–33:16, 2023.

[GJK10]   Leslie Ann Goldberg, Mark Jerrum, and Marek Karpinski. The mixing time of Glauber dynamics for coloring regular trees. *Random Structures & Algorithms*, 36(4):464–476, 2010.

[GŠV15]   Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Inapproximability for anti-ferromagnetic spin systems in the tree nonuniqueness region. *Journal of the ACM*, 62(6):Article 50, 2015.

[HS07]   Thomas P. Hayes and Alistair Sinclair. A general lower bound for mixing of single-site dynamics on graphs. *The Annals of Applied Probability*, 17(3):931–952, 2007.

[JS89]   Mark Jerrum and Alistair Sinclair. Approximating the permanent. *SIAM Journal on Computing*, 18(6):1149–1178, 1989.

[LMP09]   Brendan Lucier, Michael Molloy, and Yuval Peres. The Glauber dynamics for colourings of bounded degree trees. In *Approximation, randomization, and combinatorial optimization (RANDOM)*, volume 5687 of *Lecture Notes in Comput. Sci.*, pages 631–645, 2009.

[LP17]   David A. Levin and Yuval Peres. *Markov Chains and Mixing Times: Second Edition*. American Mathematical Society, Providence, RI, 2017.

[LS88]   Gregory F. Lawler and Alan D. Sokal. Bounds on the L^2 spectrum for Markov chains and Markov processes: a generalization of Cheeger's inequality. *Transactions of the American Mathematical Society*, 309(2):557–580, 1988.

[MSW03]   Fabio Martinelli, Alistair Sinclair, and Dror Weitz. The Ising Model on Trees: Boundary Conditions and Mixing Time. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 628–639, 2003.

[MSW04]   Fabio Martinelli, Alistair Sinclair, and Dror Weitz. Glauber dynamics on trees: boundary conditions and mixing time. *Communications in Mathematical Physics*, 250:301–334, 2004.

[RSV+14] Ricardo Restrepo, Daniel Stefankovic, Juan C. Vera, Eric Vigoda, and Linji Yang. Phase transition for Glauber dynamics for independent sets on regular trees. *SIAM Journal on Discrete Mathematics*, 28(2):835–861, 2014.

[Sin92]   Alistair Sinclair. Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combinatorics, Probability and Computing*, 1(4):351–370, 1992.

[SJ89]   Alistair Sinclair and Mark Jerrum. Approximate counting, uniform generation and rapidly mixing Markov chains. *Information and Computation*, 82(1):93–133, 1989.

[Sly10]   Allan Sly. Computational transition at the uniqueness threshold. In *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 287–296, 2010.

[SZ17]   Allan Sly and Yumeng Zhang. The Glauber dynamics of colorings on trees is rapidly mixing throughout the nonreconstruction regime. *The Annals of Applied Probability*, 27(5):2646–2674, 2017.

[TVVY12] Prasad Tetali, Juan C. Vera, Eric Vigoda, and Linji Yang. Phase transition for the mixing time of the Glauber dynamics for coloring regular trees. *The Annals of Applied Probability*, 22(6):2210–2239, 2012.

[Vig00]   Eric Vigoda. Improved bounds for sampling colorings. *Journal of Mathematical Physics*, 41(3):1555–1569, 2000.

[WZZ24]   Yulin Wang, Chihao Zhang, and Zihan Zhang. Sampling proper colorings on line graphs using $(1 + o(1))\Delta$ colors. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing (STOC)*, 2024.

A The Δn lower bound for relaxation time

In this section, we prove Theorem 5. Let $\mathbb{T} = (V, E)$ be a tree of n edges with maximum degree Δ. Let $q \geq \Delta + 1$ and let μ be the uniform distribution over all the q-edge-colorings on $\mathbb{T}$. We will give a lower bound on the relaxation time $T_{\mathrm{rel}}(\mu)$ of the Glauber dynamics on μ.

Lemma 38. *If there is an edge $\{u, v\} = e \in E$ such that $\deg_{\mathbb{T}}(u) = \deg_{\mathbb{T}}(v) = \Delta$, then for $\Delta + 1 \leq q \leq 2\Delta$, it holds that $T_{rel}(\mu) \geq \frac{\Delta n}{2(q-\Delta)^2}$.*

To prove Theorem 5, it is sufficient for us to prove Lemma 38. We will prove the lower bound in Lemma 38 by using the notion of conductance. Let Ω be the support of μ and let $P \in \mathbb{R}^{\Omega \times \Omega}$ be the transition matrix for the Glauber dynamics on μ. The conductance is defined as follow.

Definition 39. For a set $S \subseteq \Omega$, the *conductance* $\Phi(S)$ *on* S is defined as

$$\Phi(S) := \frac{\sum_{x \in S, y \in \Omega \setminus S} \mu(x) P(x, y)}{\sum_{z \in S} \mu(z)}.$$

Then the conductance $\Phi_\star$ for the whole chain is defined as $\Phi_\star := \min_{S: \mu(S) \leq 1/2} \Phi(S)$.

For convenience, we may use $\mu(S) = \sum_{x \in S} \mu(x)$ to simplify the notation.

The relaxation time is closely related to the conductance, which is known as Cheeger's inequality [SJ89, LS88]. Here, for convenience, we use the version stated in the textbook.

Lemma 40 ([LP17, Theorem 13.10]). *It holds that $\frac{\Phi_\star^2}{2} \leq T_{rel}^{-1} \leq 2\Phi_\star$.*

How, we are ready to prove Lemma 38.

Proof of Lemma 38. Let $S \subseteq \Omega$ be defined as

$$S := \{\sigma \in \Omega \mid \sigma_e = 1\}. \tag{43}$$

It is easy to see that $\mu(S) = 1/q \leq 1/2$ according to the definition. Then, in order to escape from S, the Glauber dynamics needs to update the edge e to some other color $c \neq 1$. This means

$$\mu(S)\Phi(S) = \sum_{\sigma \in S, \tau \in \Omega \setminus S} \mu(\sigma)P(\sigma, \tau) = \frac{1}{n} \sum_{\sigma \in S} \mu(\sigma) \frac{|\mathcal{L}_e^\sigma| - 1}{|\mathcal{L}_e^\sigma|} \leq \frac{1}{n} \sum_{\sigma \in S} \mu(\sigma) \mathbf{1}[|\mathcal{L}_e^\sigma| \geq 2],$$

where we use $\mathcal{L}_e^\sigma = [q] \setminus \{\sigma_h \mid h \text{ incidents } e\}$ to denote the set of available colors of e. Hence,

$$\Phi(S) \leq \frac{1}{n} \mathbf{Pr}_{\sigma \sim \mu^{e1}} \left[|\mathcal{L}_e^\sigma| \geq 2 \right]. \tag{44}$$

According to our assumption in Lemma 38, $e = \{u, v\}$ such that $\deg_{\mathbb{T}}(u) = \deg_{\mathbb{T}}(v) = \Delta$. In the rest part of the proof, we assume $\sigma \sim \mu^{e1}$. Let $\sigma(\delta_v)$ and $\sigma(\delta_u)$ be the set of colors that appears around vertex v and u, respectively. Then $|\mathcal{L}_e^\sigma| \leq 1$ is equivalent to $[q] \setminus \sigma(\delta_v) \subseteq \sigma(\delta_u)$ and

$$\mathbf{Pr}\left[|\mathcal{L}_e^\sigma| \leq 1\right] = \mathbf{Pr}\left[[q] \setminus \sigma(\delta_v) \subseteq \sigma(\delta_u)\right] = \mathbb{E}\left[\mathbf{Pr}\left[[q] \setminus \sigma(\delta_v) \subseteq \sigma(\delta_u) \mid \sigma(\delta_v)\right]\right]$$

$$= \mathbb{E}\left[\binom{q - (q - \Delta)}{\Delta - (q - \Delta)} \bigg/ \binom{q - 1}{\Delta - 1}\right] = \frac{\Delta!(\Delta - 1)!}{(2\Delta - q)!(q - 1)!}.$$

Let $q = \Delta + c$ such that $c \in [1, \Delta]$, we have

$$\mathbf{Pr}\left[|\mathcal{L}_e^\sigma| \leq 1\right] = \frac{\Delta(\Delta-1)\cdots(\Delta-c+1)}{(\Delta+c-1)(\Delta+c-2)\cdots\Delta} = \prod_{i=1}^{c} \frac{\Delta+1-i}{\Delta+c-i} \geq \left(1 - \frac{c}{\Delta}\right)^c \geq 1 - \frac{c^2}{\Delta}, \qquad (45)$$

where in the last inequality we use the Bernoulli inequality: $(1+x)^r \geq 1 + rx$ for $x \geq -1$ and $r \geq 1$. Combining (44) and (45), we have

$$\Phi_\star \leq \Phi(S) \leq \frac{c^2}{n\Delta}.$$

According to Lemma 40, we have $T_{\text{rel}} \geq \frac{n\Delta}{2c^2} = \frac{n\Delta}{2(q-\Delta)^2}$. $\qquad\square$

B Relaxation time of Glauber dynamics when $q = \Delta + 1$

In this section, we prove Theorem 4. The proof will follow the same high-level plan as Theorem 1. The crucial part is verifying the conditions required in Theorem 11. However, different from Section 3.1 where we use Theorem 11 for $\ell = \Theta(\Delta^2 \log^2 \Delta)$, in this section, we will focus on the $\ell = 1$ case. Formally, we will prove the following result.

Lemma 41. *Let $\Delta \geq 2, q = \Delta + 1$, and $\ell = 1$. The constants in Theorem 11 exist with $\alpha_0 = 4\Delta, \alpha_1 = 8$, and $\gamma = 6$.*

The bounds on $\boldsymbol{\alpha}$ and γ in Lemma 41 will be established in Appendix B.1 and Appendix B.2, respectively. We remark that we want Theorem 4 to serve as an asymptotic tight upper bound for trees with large degree, which should match the lower bound stated in Theorem 5. This means we need very accurate bounds for $\boldsymbol{\alpha}$ and γ rather than just say that they are some function of q and Δ.

How, we are able to prove Theorem 4.

Proof of Theorem 4. Let $\mathbb{T}_\ell$ be a complete Δ-regular tree with depth ℓ. Suppose $\mathbb{T}_\ell$ has n vertices, as it is a complete Δ-regular tree, it holds that $\ell = \Theta(\log_\Delta n)$. Let $\mu = \mu_\ell$ be the uniform distribution over all the $(\Delta+1)$-edge-colorings on $\mathbb{T}_\ell$. According to Lemma 41 and Theorem 11, it holds that μ satisfies approximate tensorization of variance with constant $C \leq 10\Delta \cdot 8^{\ell+1} = \Delta \cdot n^{O(1/\log\Delta)}$. This means T_{rel} is bounded by $\Delta \cdot n^{1+O(1/\log\Delta)}$. Finally, according to Corollary 8 and (4), it holds that T_{mix} is bounded by $O(1) \times T_{\text{rel}} \times \ell(\Delta \log q) \times \log\log \frac{1}{\mu_{\min}} = (\Delta \log \Delta)^2 \cdot n^{1+O(1/\log\Delta)}$. $\qquad\square$

B.1 Root tensorization on a star

In this section, we will give an upper bound for the vector $\boldsymbol{\alpha}$ in Lemma 41. Before we start, let us first recall the definition of $\mathbb{T}_\ell^\star$ and $\mu_\ell^\star$ in Definition 10. Let $\mathbb{T}_\ell^\star$ be the complete d-ary tree of depth ℓ with a hanging root edge r. For convenience, let $\widetilde{\mathbb{T}} := \mathbb{T}^\star - r$. Let $\mu = \mu_\ell^\star$ be the list coloring distribution on $\mathbb{T}^\star$, where $\mathcal{L}_r = \{1, 2\}$ and $\mathcal{L}_h = [q], q = \Delta + 1$ for other edges $h \in \mathbb{T}_\ell^\star$. In order to bound the vector $\boldsymbol{\alpha}$, we need to establish the **root**-tensorization of variance for $\mu_\ell^\star$. That is, for any function $f : \Omega(\mu) \to \mathbb{R}$, we want the following inequality to hold:

$$\mathbf{Var}_\mu\left[\mu_{\widetilde{\mathbb{T}}}[f]\right] \leq \sum_{i=0}^{\ell} \alpha_i \sum_{h \in L_i(\mathbb{T}_\ell^\star)} \mu[\mathbf{Var}_h[f]].$$

To achieve this, we do a similar canonical path analysis as we have done in Section 3.2. The different part is that we use the Cauchy's inequality in a slightly different way to make our bound

on α tighter. For each $\sigma \in \Omega(\mu^{r1})$, we pick the unique destination $\tau(\sigma) \in \Omega(\mu^{r2})$ such that $\tau(\sigma)$ is obtained by replacing the maximal $(1,2)$-alternating path from the root r in σ with an $(2,1)$-alternating path. Then, we can bound α as follows.

For convenience, given a configuration $\rho \in \Omega(\mu)$ and an edge h, we will use $\rho \oslash h$ to denote the configuration that differs at h comparing to ρ; we note that since $q = \Delta + 1$, there is at most one such configuration. If there is no such configuration, then we let $\rho \oslash h$ be ρ.

Lemma 42. *Suppose for each $\sigma \in \Omega(\mu^{r1})$, there is a path of configuration $\gamma^{\sigma} = (\gamma_0^{\sigma}, \gamma_1^{\sigma}, \cdots, \gamma_{m(\sigma)}^{\sigma})$ of length $m(\sigma)$ that $\gamma_0^{\sigma} = \sigma$, $\gamma_{m(\sigma)}^{\sigma} = \tau(\sigma)$ and $\gamma_i^{\sigma} \oplus \gamma_{i-1}^{\sigma} = \{h_i^{\sigma}\}$ for $1 \leq i \leq m(\sigma)$. Note that we use $\sigma \oplus \tau := \{h : \sigma_h \neq \tau_h\}$ to denote the symmetric difference between two configurations. Then, $\forall 0 \leq t \leq \ell$,*

$$\alpha_t \leq 2(\ell+1) \cdot \mathbb{E}_{\sigma \sim \mu^{r1}}\left[|\{i \mid h_i^{\sigma} \in L_\ell\}|\right] \cdot \max_{h \in L_t, \rho \in \Omega(\mu)} |\{\sigma \mid (\rho, \rho \oslash h) \in \gamma^{\sigma}\}|. \tag{46}$$

Proof of Lemma 42. We note that by definition

$$\mathbf{Var}_{\mu}\left[\mu_{\widetilde{\mathbb{T}}}[f]\right] = \mu_r(1)\mu_r(2)\left(\mu^{r1}(f) - \mu^{r2}(f)\right)^2$$

$$= \mu_r(1)\mu_r(2)\left(\mathbb{E}_{\sigma \sim \mu^{r1}}\left[f(\sigma) - f(\tau(\sigma))\right]\right)^2, \tag{47}$$

Then, we could use the telescoping sum for $f(\sigma) - f(\tau(\sigma))$ and we have

$$f(\sigma) - f(\tau(\sigma)) = \sum_{i=1}^{m(\sigma)} \left(f(\gamma_{i-1}^{\sigma}) - f(\gamma_i^{\sigma})\right). \tag{48}$$

Now, combining (47) and (48), we have

$$\left(\mathbb{E}_{\sigma \sim \mu^{r1}}\left[f(\sigma) - f(\tau(\sigma))\right]\right)^2 = \left(\mathbb{E}_{\sigma \sim \mu^{r1}}\left[\sum_{i=1}^{m(\sigma)} \left(f(\gamma_{i-1}^{\sigma}) - f(\gamma_i^{\sigma})\right)\right]\right)^2$$

$$= \left(\sum_{t=0}^{\ell} \mathbb{E}_{\sigma \sim \mu^{r1}}\left[\sum_{i:h_i^{\sigma} \in L_t} \left(f(\gamma_{i-1}^{\sigma}) - f(\gamma_i^{\sigma})\right)\right]\right)^2,$$

where L_t is the set of edges at level t of the tree $\mathbb{T}$. By using the Cauchy's inequality, we have

$$\left(\mathbb{E}_{\sigma \sim \mu^{r1}}\left[f(\sigma) - f(\tau(\sigma))\right]\right)^2 \leq (\ell+1)\sum_{t=0}^{\ell}\left(\mathbb{E}_{\sigma \sim \mu^{r1}}\left[\sum_{i:h_i^{\sigma} \in L_t} \left(f(\gamma_{i-1}^{\sigma}) - f(\gamma_i^{\sigma})\right)\right]\right)^2. \tag{49}$$

For convenience, for $0 \leq t \leq \ell$, we let

$$\mathbf{Var}_t := \left(\mathbb{E}_{\sigma \sim \mu^{r1}}\left[\sum_{i:h_i^{\sigma} \in L_t} \left(f(\gamma_{i-1}^{\sigma}) - f(\gamma_i^{\sigma})\right)\right]\right)^2. \tag{50}$$

Since μ^{r1} is a uniform distribution, we let $p^{r1} := 1/\left|\Omega(\mu^{r1})\right|$. Now, applying Cauchy's inequality

again on $\mathbf{Var}_t$, we have

$$\mathbf{Var}_t = (p^{r1})^2 \left(\sum_{\sigma \in \Omega(\mu^{r1})} \sum_{i:h_i^\sigma \in L_t} \left(f(\gamma_{i-1}^\sigma) - f(\gamma_i^\sigma) \right) \right)^2$$

$$\leq p^{r1} \left(\sum_{\sigma \in \Omega(\mu^{r1})} |\{i \mid h_i^\sigma \in L_t\}| \right) \cdot p^{r1} \left(\sum_{\sigma \in \Omega(\mu^{r1})} \sum_{i:h_i^\sigma \in L_t} \left(f(\gamma_{i-1}^\sigma) - f(\gamma_i^\sigma) \right)^2 \right)$$

$$= \mathbb{E}_{\sigma \sim \mu^{r1}} \left[|\{i \mid h_i^\sigma \in L_t\}| \right] \cdot p^{r1} \sum_{h \in L_t} \sum_{\rho \in \Omega(\mu)} |\{\sigma \mid (\rho, \rho \oslash h) \in \gamma^\sigma\}| \left(f(\rho) - f(\rho \oslash h) \right)^2$$

$$\leq \mathbb{E}_{\sigma \sim \mu^{r1}} \left[|\{i \mid h_i^\sigma \in L_t\}| \right] \cdot \max_{h,\rho} |\{\sigma \mid (\rho, \rho \oslash h) \in \gamma^\sigma\}| \cdot p^{r1} \sum_{h \in L_t} \sum_{\rho \in \Omega(\mu)} \left(f(\rho) - f(\rho \oslash h) \right)^2$$

$$= \mathbb{E}_{\sigma \sim \mu^{r1}} \left[|\{i \mid h_i^\sigma \in L_t\}| \right] \cdot \max_{h,\rho} |\{\sigma \mid (\rho, \rho \oslash h) \in \gamma^\sigma\}| \cdot 8 \sum_{h \in L_t} \mu[\mathbf{Var}_h[f]] \qquad (51)$$

Then (46) could be proved by combining (47), (49), (50), and (51), where we use the fact that $\mu_r(1) = \mu_r(2) = 1/2$. $\qquad \square$

Now, we are ready to bound α in Lemma 41.

Proof (bound α in Lemma 41). Recall that we only consider $\ell = 1$ in Lemma 41. This choice makes everything simple and we could finish the proof by a case study.

For convenience, given a configuration $\rho \in \Omega(\mu)$, we will use AP_ρ to denote the maximal $(1,2)$-alternating path or $(2,1)$-alternating path in ρ initiated from the root edge r depending on ρ_r. We construct γ^σ by considering two cases according to the length of AP_σ: (1) $|\mathrm{AP}_\sigma| = 1$; (2) $|\mathrm{AP}_\sigma| = 2$.

When $|\mathrm{AP}_\sigma| = 1$, it holds that $\mathrm{AP}_\sigma = \{r\}$ and we have $\tau(\sigma) = \sigma \oslash r$.

When $|\mathrm{AP}_\sigma| = 2$, it holds that $\mathrm{AP}_\sigma = \{r, h\}$ where h is some edge in the first level. Then we can route σ to τ by noticing that $\tau = \sigma \oslash h \oslash r \oslash h$.

First, we claim that α_1 is bounded by 8. Since for $\rho \in \Omega(\mu)$ and $h \in L_1$, there is only 1 possible starting point σ. When $|\mathrm{AP}_\rho| = 1$, we know that $\rho \oslash h = \tau$, so that we could determine σ unqiuely. When $|\mathrm{AP}_\rho| = 2$, we know that $\rho = \sigma$. Formally, for every $\rho \in \Omega(\mu)$ and $h \in L_1$, it holds that

$$|\{\sigma \mid (\rho, \rho \oslash h) \in \gamma^\sigma\}| = 1$$

Then, by our construction of γ^σ, it is straight forward to see that

$$\mathbb{E}_{\sigma \sim \mu^{r1}} \left[|\{i \mid h_i^\sigma \in L_1\}| \right] \leq 2\mathbf{Pr}\left[|\mathrm{AP}_\sigma| = 2 \right] \leq 2.$$

Now, our claim for α_1 comes from Lemma 42

Then, we claim that α_0 is bounded by 4Δ. Given $\rho \in \Omega(\mu)$ and $h = r \in L_0$, we will consider two cases: (1) $|\mathrm{AP}_\rho| = 1$; (2) $|\mathrm{AP}_\rho| = 2$. When $|\mathrm{AP}_\rho| = 1$, there are at most Δ possible starting configuration σ: (1) $\sigma = \rho$; (2) $\sigma = \rho \oslash g$ for some $g \in L_1$. This means

$$|\{\sigma \mid (\rho, \rho \oslash h) \in \gamma_\sigma\}| \leq \Delta.$$

When $|\mathrm{AP}_\rho| = 2$, we know that $|\{\sigma \mid (\rho, \rho \oslash h) \in \gamma_\sigma\}| = 0$ from our construction of γ_σ. Also, we know that

$$\mathbb{E}_{\rho \sim \mu^{r1}} \left[|\{i \mid h_i^\sigma \in L_1\}| \right] = 1.$$

Now, our claim for α_0 comes from Lemma 42. $\qquad \square$

B.2 Approximate tensorization on a star

In this section, we bound γ in Lemma 41. Let $S_\Delta = (V, [\Delta])$ be the star graph with Δ edges. In the rest part of this section, we fix $q = q(\Delta) = \Delta + 1$ and consider the uniform distribution μ_Δ over all the q-edge-colorings on S_Δ. We will prove approximate tensorization of variance for μ_Δ which is formally stated as in Lemma 43. Then, Lemma 43 directly implies that $\gamma \leq C = \exp(\pi^2/6) \leq 6$.

Lemma 43. *For $\Delta \geq 1$, μ_Δ satisfies approximate tensorization of variance with constant $C = \exp(\pi^2/6)$.*

In the rest part of this section, we will focus on proving Lemma 43. The proof relies on the standard local-to-global argument for the high-dimensional expander [AL20].

Before we start the proof, we first give a brief introduction on this technique. For simplicity we specialize the notions for the special distribution μ that we will consider in this section.

For any integer $\Delta \geq 2$, let $\mu = \mu_\Delta$, we consider the following matrices:

- Let $\Psi^{\mathrm{cor}} \in \mathbb{R}^{q\Delta \times q\Delta}$ be the correlation matrix of μ defined as follow

$$\forall u, v \in [\Delta], \quad \forall a, b \in [q], \quad \Psi^{\mathrm{cor}}(ua, vb) := \mu_v^{ua}(b) - \mu_v(b). \tag{52}$$

- Let $P_\Delta \in \mathbb{R}^{q\Delta \times q\Delta}$ be the transition matrix of the non-lazy local random walk defined as follow

$$\forall u, v \in [\Delta], \quad \forall a, b \in [q], \quad P_\Delta(ua, vb) := \frac{\mathbf{1}[u \neq v]}{\Delta - 1} \cdot \mu_v^{ua}(b). \tag{53}$$

Lemma 44 ([AL20, Theorem 1.5]). *The distribution μ_Δ satisfies approximate tensorization of variance with constant $C = \prod_{i=0}^{\Delta-2}(1 - \lambda_2(P_{\Delta-i}))^{-1}$ where λ_2 refers to the second largest eigenvalue of a matrix.*

According to the definition of μ, (52) and (53), it is straightforward to verify that

$$\Psi_\Delta^{\mathrm{cor}} = (\Delta - 1)P_\Delta - \mathbf{1}\mathbf{1}^\top/q + \mathbb{I}, \tag{54}$$

where we use $\mathbf{1}$ to denote the all-1 vector over $\mathbb{R}^{q\Delta}$ and $\mathbb{I}$ is the identity matrix over $\mathbb{R}^{q\Delta \times q\Delta}$. Since $\Psi_\Delta^{\mathrm{cor}}\mathbf{1} = 0$, according to (54), we have $\lambda_{\max}(\Psi_\Delta^{\mathrm{cor}}) = (\Delta - 1)\lambda_2(P_\Delta) + 1$, which means

$$\lambda_2(P_\Delta) = (\lambda_{\max}(\Psi_\Delta^{\mathrm{cor}}) - 1)/(\Delta - 1). \tag{55}$$

We have the following upper bound for $\lambda_{\max}(\Psi_\Delta^{\mathrm{cor}})$.

Lemma 45. *It holds that $\lambda_{\max}(\Psi_\Delta^{\mathrm{cor}}) \leq 1 + \frac{1}{\Delta}$.*

Now, we are ready to prove Lemma 43.

proof of Lemma 43. According to Lemma 44, (55), and Lemma 45, it holds that

$$C \leq \exp\left(\sum_{i=0}^{\Delta-2} \frac{1}{(\Delta - i)(\Delta - i - 1)}\right) \leq \exp\left(\sum_{i=1}^{\infty} \frac{1}{i^2}\right) = \exp(\pi^2/6). \qquad \square$$

We only left to prove Lemma 45.

Proof of Lemma 45. According to the definition of Ψ^{cor} in (52), for $u, v \in S_\Delta$ and $a, b \in [q]$, we have

$$\Psi_\Delta^{\text{cor}}(ua, vb) = -\frac{1}{\Delta + 1} + \mathbf{1}[u \neq v \wedge a \neq b]\frac{1}{\Delta} + \mathbf{1}[u = v \wedge a = b]. \tag{56}$$

We may consider Ψ^{cor} as a block matrix by letting coordinates with the same color in the same block. Then, it holds that

$$\Psi_\Delta^{\text{cor}} = \left(1 + \frac{1}{\Delta}\right) \mathbb{I} + \left(\frac{1}{\Delta} - \frac{1}{\Delta + 1}\right) \mathbb{1}\mathbb{1}^\top - \frac{1}{\Delta}\left(\mathbb{I}_q \otimes \left(\mathbb{1}_\Delta \mathbb{1}_\Delta^\top\right) + \left(\mathbb{1}_q \otimes \mathbb{I}_\Delta\right)\left(\mathbb{1}_q \otimes \mathbb{I}_\Delta\right)^\top\right), \tag{57}$$

where $\otimes$ is the Kronecker product and we let $\mathbb{I}$ denote the identity matrix over $\mathbb{R}^{q\Delta \times q\Delta}$, $\mathbb{1}$ denote the all-1 vector over $\mathbb{R}^{q\Delta}$. Similarly, for $1 \leq t \leq q\Delta$, we let $\mathbb{I}_t$ be the identity matrix in $\mathbb{R}^{t \times t}$ and we let $\mathbb{1}_t$ be the all-1 vector of length t.

According to (52), it holds that $\Psi_\Delta^{\text{cor}}\mathbb{1} = 0$. Since Ψ_Δ^{cor} is a symmetric matrix, to prove Lemma 45, we only need to bound $\frac{f^\top \Psi_\Delta^{\text{cor}} f}{f^\top f}$ for $f \neq 0$ and $f \perp \mathbb{1}$. By (57), we have

$$f^\top \Psi_\Delta^{\text{cor}} f \overset{(\star)}{\leq} f^\top \left(\left(1 + \frac{1}{\Delta}\right)\mathbb{I} + \left(\frac{1}{\Delta} - \frac{1}{\Delta + 1}\right)\mathbb{1}\mathbb{1}^\top\right) f \overset{f \perp \mathbb{1}}{=} \left(1 + \frac{1}{\Delta}\right) f^\top f,$$

where $(\star)$ holds by

$$\mathbb{I}_q \otimes \left(\mathbb{1}_\Delta \mathbb{1}_\Delta^\top\right) \succcurlyeq 0 \quad \text{and} \quad \left(\mathbb{1}_q \otimes \mathbb{I}_\Delta\right)\left(\mathbb{1}_q \otimes \mathbb{I}_\Delta\right)^\top \succcurlyeq 0. \qquad \square$$

C Block factorization via root-factorization: Proof of Theorem 21

In this section, we prove Theorem 21. Note that Theorem 21 is a generalization of Theorem 11, and the proof presented here is a generalization of the proof of Theorem 11 presented in Section 4.

Recall the definitions of $\mathbb{T}_k, \mathbb{T}_k^\star$ and $\mu_k, \mu_k^\star, \mu_k^{\star,1}$ from Definitions 9 and 10. We will prove a more general version of Theorem 21, which will yield Theorem 21 as a corollary.

Definition 46. For $\ell \geq 1$, let $\mathcal{A}$ be a collection of edge sets of $\mathbb{T}_\ell^\star$ such that $\forall S \in \mathcal{A}$ satisfies:

- $|S| \geq 2$, i.e., S is not a singleton edge in $\mathbb{T}_\ell^\star$;

- $S \cap L_\ell(\mathbb{T}_\ell^\star) = \varnothing$.

Then, we define the collection $\mathcal{B}$ as

$$\mathcal{B} := \{\{e\} \mid e \in \mathbb{T}_\ell^\star\} \cup \mathcal{A}, \tag{58}$$

so that (58) is a generalization of (17). Then, following (18), we say that *approximate **root**-factorization of variance* for $\mu_\ell^\star$ holds with constants $C(B)$ if for all $f : \Omega(\mu_\ell^\star) \to \mathbb{R}$

$$\mathbf{Var}_{\mu_\ell^\star}\left[\mu_{\mathbb{T}_\ell^\star \setminus \{r\}}[f]\right] \leq \sum_{B \in \mathcal{B}} C(B)\mu_\ell^\star(\mathbf{Var}_B[f]).$$

Lemma 47. *Let $\Delta \geq 2$, $q \geq \Delta + 1$ be two integers. Suppose there exists an integer $\ell \geq 1$ where the following holds:*

- *There exist constants $\alpha = (\alpha_0, \ldots, \alpha_\ell)$ and β and approximate **root**-factorization of variance for $\mu_k^\star$ holds with constants $C(\{e\}) = \alpha_i$ where $e \in L_i(\mathbb{T}_\ell^\star)$ and $C(B) = \beta$ where $|B| \geq 2$;*

- *There exists a constant γ such that for $1 \leq j \leq \ell$, both μ_j and $\mu_j^{\star,1}$ satisfy approximate tensorization of variance with constant γ.*

Then, for any $k \geq 1$, μ_k satisfies the following block factorization of variance: for any $f : \Omega(\mu_k) \to \mathbb{R}$,

$$\mathbf{Var}_{\mu_k}[f] \leq \left(\sum_{t=1}^{k} F_k(t) \sum_{e \in L_t(\mathbb{T}_k)} \mu[\mathbf{Var}_e[f]] \right) + \beta \sum_{\substack{1 \leq t < k \\ t \equiv k \pmod{\ell}}} F_t(t) \sum_{r \in L_t(\mathbb{T}_k)} \sum_{B \in \mathcal{A}^r} \mu[\mathbf{Var}_B[f]], \qquad (59)$$

where $F_{\cdot}(\cdot)$ is the function defined in (20) and for every edge $r \leq k - \ell$, we let $\mathcal{A}^r$ be a copy of $\mathcal{A}$ by seeing the sub-tree rooted at r of depth ℓ as $\mathbb{T}_\ell^\star$.

Now we can prove Theorem 21 as a direct corollary of Lemma 47.

Proof of Theorem 21. We set $\mathcal{B}$ as in (17). Since Theorem 21 assumes $\ell \geq 2$, we know this specific $\mathcal{B}$ satisfies the requirements in Definition 46 and (58). Now, according to Lemma 47, we know that we have the block factorization in (59). Furthermore, according to Section 4.3, we know that the function $F_{\cdot}(\cdot)$ can be bounded by $\widehat{F}_k$ defined in (32), which means for any $1 \leq t \leq k$, we have

$$F_k(t) \leq \gamma \cdot \left(\max_{0 \leq j \leq \ell} \alpha_j \cdot \sum_{i=0}^{\lfloor k/\ell \rfloor} \alpha_\ell^i + \max\{1, \alpha_0\} \right).$$

We finish the prove by comparing terms in (59) and the block factorization claimed in Theorem 21 using the fact that variance is non-negative. $\qquad \square$

In the rest part of this section, we will prove Lemma 47. Fix parameters d, q, ℓ and $\boldsymbol{\alpha}$ in the theorem. Our proof is based on an induction of the depth k.

Base case ($k \leq \ell$): When $k \leq \ell$, (59) follows from the second assumption of Lemma 47 with $F_k(t) = \gamma$.

Inductive case ($k > \ell$): Suppose (59) holds for all smaller $k' < k$. We prove that the theorem holds for k. For convenience, we denote $\mu = \mu_k$ and $\mathbb{T} = \mathbb{T}_k$. Let $R = \cup_{i=k-\ell+1}^{k} L_i(\mathbb{T})$ be the edge set that contains all the ℓ levels in the bottom of $\mathbb{T}$. According to the law of total variance,

$$\mathbf{Var}_\mu[f] \stackrel{(21)}{=} \mu[\mathbf{Var}_R[f]] + \mathbf{Var}_\mu[\mu_R[f]]. \qquad (60)$$

The first term in the RHS of (60) could be bounded by Claim 27.

Next, we bound the second term in the RHS of (60) by induction hypothesis. Note that for any $\sigma \in \Omega(\mu)$, the value of function $\mu_R[f](\sigma)$ depends only on $\sigma_{\mathbb{T} \setminus R}$. By viewing $\mu_R[f]$ as a function $f' : \Omega(\mu_{\mathbb{T} \setminus R}) \to \mathbb{R}$, we have $\mathbf{Var}_\mu[\mu_R[f]] = \mathbf{Var}_{\mu_{\mathbb{T}-R}}[f'] = \mathbf{Var}_{\mu_{k-\ell}}[f']$, where the last equation uses the fact that the marginal distribution $\mu_{\mathbb{T} \setminus R}$ is the same as $\mu_{k-\ell}$ and $\mu_{k-\ell}$ is the uniform distribution

over proper q-edge-coloring in tree $\mathbb{T}_{k-\ell}$. By the I.H. on $\mu_{k-\ell} = \mu_{\mathbb{T}\backslash R}$, we have

$$\mathbf{Var}_\mu\left[\mu_R[f]\right] \leq \sum_{t=1}^{k-\ell} F_{k-\ell}(t) \sum_{h\in L_t(\mathbb{T})} \mu_{\mathbb{T}\backslash R}[\mathbf{Var}_h\left[f'\right]] + \beta \sum_{\substack{1\leq t<k-\ell \\ t\equiv k(\mathrm{mod}\ell)}} F_t(t) \sum_{r\in L_t(\mathbb{T})} \sum_{B\in\mathcal{A}^r} \mu_{\mathbb{T}\backslash R}[\mathbf{Var}_B\left[f'\right]]$$

$$\overset{(*)}{=} \sum_{t=1}^{k-\ell-1} F_{k-\ell}(t) \sum_{h\in L_t(\mathbb{T})} \mu[\mathbf{Var}_h\left[\mu_R[f]\right]] + F_{k-\ell}(k-\ell) \sum_{h\in L_{k-\ell}(\mathbb{T})} \mu\left[\mathbf{Var}_{\mathbb{T}^h}\left[\mu_R[f]\right]\right]$$

$$+ \beta \sum_{\substack{1\leq t<k-\ell \\ t\equiv k(\mathrm{mod}\ell)}} F_t(t) \sum_{r\in L_t(\mathbb{T})} \sum_{B\in\mathcal{A}^r} \mu[\mathbf{Var}_B\left[\mu_R[f]\right]]$$

$$\overset{(24)}{\leq} \sum_{t=1}^{k-\ell-1} F_{k-\ell}(t) \sum_{h\in L_t(\mathbb{T})} \mu[\mathbf{Var}_h\left[f\right]] + F_{k-\ell}(k-\ell) \sum_{h\in L_{k-\ell}(\mathbb{T})} \mu\left[\mathbf{Var}_{\mathbb{T}^h}\left[\mu_{\widetilde{\mathbb{T}}^h}[f]\right]\right] \qquad (61)$$

$$+ \beta \sum_{\substack{1\leq t<k-\ell \\ t\equiv k(\mathrm{mod}\ell)}} F_t(t) \sum_{r\in L_t(\mathbb{T})} \sum_{B\in\mathcal{A}^r} \mu[\mathbf{Var}_B\left[f\right]],$$

where we use $\mathbb{T}^h$ to denote the subtree with the hanging root edge h and use $\widetilde{\mathbb{T}}^h := \mathbb{T}^h \setminus h$ to denote the subtree beneath the edge h. The equation $(*)$ holds because

- for all $h \in L_t(\mathbb{T})$ with $t < k - \ell$, all the neighboring edges of h are in $\mathbb{T} \setminus R$, i.e. $\partial h \subseteq \mathbb{T} \setminus R$, by conditional independence, $\mu_{\mathbb{T}\backslash R}[\mathbf{Var}_h\left[f'\right]] = \mu[\mathbf{Var}_h\left[\mu_R[f]\right]]$;

- for all $1 \leq t < k - \ell$ such that $t \equiv k(\mathrm{mod}\ell)$, we know that $t \leq k - 2\ell$, and for $r \in L_t(\mathbb{T}_k)$, by the definition of $\mathcal{A}^r$ and $\mathcal{A}$ in Definition 46, we know that $\forall B \in \mathcal{A}^r$, $\bar{\partial}B \subseteq \mathbb{T} \setminus R$, which, by conditional independence, implies $\mu_{\mathbb{T}\backslash R}[\mathbf{Var}_B\left[f'\right]] = \mu[\mathbf{Var}_B\left[\mu_R[f]\right]]$;

- for all $h \in L_{k-\ell}(\mathbb{T})$, $\mu_{\mathbb{T}\backslash R}[\mathbf{Var}_h\left[f'\right]] = \mu\left[\mathbf{Var}_{\mathbb{T}^h}\left[\mu_R[f]\right]\right]$ because for any $\sigma \in \Omega(\mu)$, the value of $\mu_R[f](\sigma)$ is independent from $\sigma_{\mathbb{T}_h\backslash h}$.

Since $\mu\left[\mathbf{Var}_{\mathbb{T}^h}\left[\mu_{\widetilde{\mathbb{T}}^h}[f]\right]\right] = \sum_{\sigma\in\Omega(\mu_{\mathbb{T}\backslash\widetilde{\mathbb{T}}^h})} \mu_{\mathbb{T}\backslash\widetilde{\mathbb{T}}^h}(\sigma)\mathbf{Var}_{\mu^\sigma}\left[\mu_{\widetilde{\mathbb{T}}^h}[f]\right]$. Fix one σ. One can view f as a function from $\Omega(\mu_{\widetilde{\mathbb{T}}^h}^\sigma)$ to $\mathbb{R}$. By applying the first assumption of Lemma 47 to each σ individually and rearranging,

$$\sum_{h\in L_{k-\ell}(\mathbb{T})} \mu\left[\mathbf{Var}_{\mathbb{T}^h}\left[\mu_{\widetilde{\mathbb{T}}^h}[f]\right]\right] \leq \sum_{i=0}^{\ell} \alpha_i \sum_{g\in L_{k-\ell+i}(\mathbb{T})} \mu[\mathbf{Var}_g\left[f\right]] + \beta \sum_{r\in L_{k-\ell}(\mathbb{T})} \sum_{B\in\mathcal{A}^r} \mu[\mathbf{Var}_B\left[f\right]]. \qquad (62)$$

Combining (60), (26) in Claim 27, (61), and (62), it holds that

$$\mathbf{Var}_\mu\left[f\right] \leq \sum_{t=1}^{k-\ell-1} F_{k-\ell}(t) \sum_{h\in L_t(\mathbb{T})} \mu[\mathbf{Var}_h\left[f\right]]$$

$$+ F_{k-\ell}(k-\ell) \cdot \alpha_0 \sum_{h\in L_{k-\ell}(\mathbb{T})} \mu[\mathbf{Var}_h\left[f\right]]$$

$$+ \sum_{i=1}^{\ell} (\alpha_i F_{k-\ell}(k-\ell) + \gamma) \sum_{h\in L_{k-\ell+i}(\mathbb{T})} \mu[\mathbf{Var}_h\left[f\right]]$$

$$+ \beta \sum_{\substack{1\leq t<k \\ t\equiv k(\mathrm{mod}\ell)}} F_t(t) \sum_{r\in L_t(\mathbb{T})} \sum_{B\in\mathcal{A}^r} \mu[\mathbf{Var}_B\left[f\right]].$$

The above inequality immediately gives the recurrence of F as in (20) and finishes the proof.

D Monotonicity of the Glauber dynamics (Proof of Lemma 14)

In this section, we prove Lemma 14. Recall that $\mathbb{T} = (V, E)$ is a tree with root r and for $H \subseteq E$ the tree $\mathbb{T}_H = (V_H, H \subseteq E)$ is a connected sub-tree of $\mathbb{T}$ containing r. We use μ and ν to denote the distribution of uniform q-edge-coloring on $\mathbb{T}$ and $\mathbb{T}_H$, respectively.

By the definition of approximate tensorization of variance and log-Sobolev constant in Definition 6 and (3), to prove Lemma 14, it is sufficient for us to prove the following inequalities for every function $g : \Omega(\mu) \to \mathbb{R}$:

$$\mathbf{Var}_\nu [g] \leq Cq \sum_{e \in H} \nu[\mathbf{Var}_e [g]] \tag{63}$$

$$\mathbf{Ent}_\nu \left[g^2 \right] \leq c_{\text{sob}}^{-1}(\mu) q \sum_{e \in H} \nu[\mathbf{Var}_e [g]], \tag{64}$$

where (63) is for the first part of Lemma 14 and (64) is for the second part of Lemma 14. We recall the notation $\mathbf{Ent}[\cdot]$ for entropy is defined in (2).

For a fixed function $g : \Omega(\nu) \to \mathbb{R}$, we construct a function $f : \Omega(\mu) \to \mathbb{R}$ as follow

$$\forall \sigma \in \Omega(\nu), \quad f(\sigma) := g(\sigma_H), \tag{65}$$

where we use σ_H to denote the sub-vector obtained by restricting σ to the edge set H. Then (63) and (64) follow directly from the following claims. Through the whole proof, the crucial property that we use is $\nu = \mu_H$.

Claim 48. $\mathbf{Var}_\mu [f] \geq \mathbf{Var}_\nu [g]$ *and* $\mathbf{Ent}_\mu \left[f^2 \right] \geq \mathbf{Ent}_\nu \left[g^2 \right]$.

Proof. We focus on proving $\mathbf{Ent}_\mu \left[f^2 \right] \geq \mathbf{Var}_\nu \left[g^2 \right]$. The fact $\mathbf{Var}_\mu [f] \geq \mathbf{Var}_\nu [g]$ can be proved using a similar argument. Let $X \sim \mu$ be a random edge coloring and let $F = f^2(X)$ be a random variable. Since $X \sim \mu$ and $\nu = \mu_H$, we know that $X_H \sim \nu$. This means we can write

$$\mathbf{Ent}_\mu \left[f^2 \right] = \mathbf{Ent}_{X \sim \mu} \left[f^2(X) \right] = \mathbf{Ent}[F]$$

$$\mathbf{Ent}_\nu \left[g^2 \right] = \mathbf{Ent}_{X \sim \mu} \left[g^2(X_H) \right] = \mathbf{Ent}_{X \sim \mu} \left[\mathbb{E} \left[f^2(X) \mid X_H \right] \right] = \mathbf{Ent} \left[\mathbb{E} \left[F \mid X_H \right] \right], \tag{66}$$

where the second equation in (66) follows from the definition of f in (65). Finally, according to the law of total entropy, $\mathbf{Ent}[F] - \mathbf{Ent}[\mathbb{E}[F \mid X_H]] = \mathbb{E}[\mathbf{Ent}[F \mid X_H]]$, which is ≥ 0 according to the non-negativity of entropy. $\qquad\square$

Claim 49. $\sum_{e \in E} \mu[\mathbf{Var}_e [f]] \leq q \sum_{h \in H} \nu[\mathbf{Var}_h [g]]$.

Proof. For $e \notin H$, note that $\mu[\mathbf{Var}_e [f]] = 0$ by the definition of f. Then for every $h \in H$, we will

show that $\mu[\mathbf{Var}_h[f]] \leq q\nu[\mathbf{Var}_h[g]]$. By the definition of $\mu[\mathbf{Var}_h[f]]$, we have

$$
\begin{aligned}
\mu[\mathbf{Var}_e[f]] &= \frac{1}{2} \sum_{\sigma \in \Omega(\mu_{E \setminus h})} \mu_{E \setminus h}(\sigma) \sum_{a,b \in \Omega(\mu_h^\sigma)} \mu_h^\sigma(a) \mu_h^\sigma(b) (f(\sigma,a) - f(\sigma,b))^2 \\
&\leq \frac{1}{2} \sum_{\sigma \in \Omega(\mu_{E \setminus h})} \sum_{a,b \in \Omega(\mu_h^\sigma)} \mu(\sigma,a)(f(\sigma,a) - f(\sigma,b))^2 \\
&= \frac{1}{2} \sum_{\alpha \in \Omega(\mu_{H \setminus h})} \sum_{\beta \in \Omega(\mu_{E \setminus H}^\alpha)} \sum_{a,b \in \Omega(\mu_h^{\alpha,\beta})} \mu(\alpha,\beta,a)(f(\alpha,\beta,a) - f(\alpha,\beta,b))^2 \\
&= \frac{1}{2} \sum_{\alpha \in \Omega(\mu_{H \setminus h})} \mu_{H \setminus h}(\alpha) \sum_{a,b \in \Omega(\mu_h^\alpha)} \mu_h^\alpha(a) \sum_{\beta \in \Omega(\mu_{E \setminus H}^{\alpha,a}) \cap \Omega(\mu_{E \setminus H}^{\alpha,b})} \mu_{E \setminus H}^{\alpha,a}(\beta)(g(\alpha,a) - g(\alpha,b))^2 \\
&\leq \frac{1}{2} \sum_{\alpha \in \Omega(\mu_{H \setminus h})} \mu_{H \setminus h}(\alpha) \sum_{a,b \in \Omega(\mu_h^\alpha)} \mu_h^\alpha(a)(g(\alpha,a) - g(\alpha,b))^2 \\
&\leq q \cdot \frac{1}{2} \sum_{\alpha \in \Omega(\mu_{H \setminus h})} \mu_{H \setminus h}(\alpha) \sum_{a,b \in \Omega(\mu_h^\alpha)} \mu_h^\alpha(a) \mu_h^\alpha(b)(g(\alpha,a) - g(\alpha,b))^2 = q\nu[\mathbf{Var}_h[g]],
\end{aligned}
$$

where in the last inequality, we use the fact that $\mu_h^\alpha(b) \geq 1/q$ and $\nu = \mu_H$. $\qquad \square$

Proof of Lemma 14. (First part): suppose μ satisfies approximate tensorization of variance with constant C, which could be written as the following inequality: for any $f : \Omega(\mu) \to \mathbb{R}$,

$$
\mathbf{Var}_\mu[f] \leq C \sum_{e \in E} \mu[\mathbf{Var}_e[f]].
$$

Then, for any function $g : \Omega(\nu) \to \mathbb{R}$, we define the function f as in (65). According to Claim 48 and Claim 49, it hold that

$$
\mathbf{Var}_\nu[g] \leq \mathbf{Var}_\mu[f] \leq C \sum_{e \in E} \mu[\mathbf{Var}_e[f]] \leq Cq \sum_{h \in H} \nu[\mathbf{Var}_h[g]],
$$

which proves (63) and hence the first part of Lemma 14.

(Second part): Suppose μ has log-Sobolev constant $c_{\text{sob}}(\mu)$. According to definition, this means for any $f : \Omega(\mu) \to \mathbb{R}$, it holds that

$$
\mathbf{Ent}_\mu[f^2] \leq c_{\text{sob}}^{-1}(\mu) \sum_{e \in E} \mu[\mathbf{Var}_e[f]].
$$

Then, for any function $g : \Omega(\nu) \to \mathbb{R}$, we define the function f as in (65). According to Claim 48 and Claim 49, it hold that

$$
\mathbf{Ent}_\nu[g^2] \leq \mathbf{Ent}_\mu[f^2] \leq c_{\text{sob}}^{-1}(\mu) \sum_{e \in E} \mu[\mathbf{Var}_e[f]] \leq q c_{\text{sob}}^{-1}(\mu) \sum_{h \in H} \nu[\mathbf{Var}_h[g]],
$$

which gives (64) and hence proves the second part of Lemma 14. $\qquad \square$

E Monotonicity of the heat-bath edge dynamics

In this section, we prove Lemma 24. Recall that $\mathbb{T} = (V, E)$ is a tree with root r and for $H \subseteq E$ the tree $\mathbb{T}_H = (V_H, H \subseteq E)$ is a connected sub-tree of $\mathbb{T}$ containing r We use μ and ν to denote

the distribution of uniform q-edge-coloring on $\mathbb{T}$ and $\mathbb{T}_H$, respectively. The proof follows the same high level plan as the proof of Lemma 14. The only difference is that we need to consider edge pairs instead of singleton edges.

Fix a function $g : \Omega(\nu) \to \mathbb{R}$, let the function $f : \Omega(\mu) \to \mathbb{R}$ be defined as in (65). We note that Lemma 24 follows directly from Claim 48, Claim 49, and the following lemma.

Lemma 50. *Let $\mathcal{P}_E$ and $\mathcal{P}_H$ be the sets of neighboring edge pairs in E and H, respectively. It holds that $\sum_{\{e,h\}\in\mathcal{P}_E}\mu[\mathbf{Var}_{\{e,h\}}[f]] \leq q^2 \sum_{\{e,h\}\in\mathcal{P}_H}\nu[\mathbf{Var}_{\{e,h\}}[g]] + 2q^2 \sum_{e\in H}\nu[\mathbf{Var}_e[f]]$.*

Proof. To finish the proof, we do a case study for $\{e,h\} \in \mathcal{P}_E$, we claim that:

1. for $e \notin H$ and $h \notin H$, $\mu[\mathbf{Var}_{\{e,h\}}[f]] = 0$;

2. for $e \in H$ and $h \in H$, $\mu[\mathbf{Var}_{\{e,h\}}[f]] \leq q^2\nu[\mathbf{Var}_{\{e,h\}}[g]]$;

3. for $e \in H$ and $h \notin H$, $\mu[\mathbf{Var}_{\{e,h\}}[f]] \leq q\nu[\mathbf{Var}_e[f]]$.

Then, Lemma 50 follows from these claims directly. The first case is trivial due to the definition of the function f. For the second case, by definition of $\mu[\mathbf{Var}_{\{e,h\}}[f]]$, we have

$$
\begin{aligned}
\mu[\mathbf{Var}_{\{e,h\}}[f]] &= \frac{1}{2}\sum_{\sigma\in\Omega(E\backslash\{e,h\})}\mu_{E\backslash\{e,h\}}(\sigma)\sum_{a,b\in\Omega(\mu^\sigma_{e,h})}\mu^\sigma_{e,h}(a)\mu^\sigma_{e,h}(b)(f(\sigma,a)-f(\sigma,b))^2 \\
&\leq \frac{1}{2}\sum_{\sigma\in\Omega(E\backslash\{e,h\})}\sum_{a,b\in\Omega(\mu^\sigma_{e,h})}\mu(\sigma,a)(f(\sigma,a)-f(\sigma,b))^2 \\
&= \frac{1}{2}\sum_{\alpha\in\Omega(\mu_{H\backslash\{e,h\}})}\sum_{\beta\in\Omega(\mu^\sigma_{E\backslash H})}\sum_{a,b\in\Omega(\mu^{\alpha,\beta}_{e,f})}\mu(\alpha,\beta,a)(f(\alpha,\beta,a)-f(\alpha,\beta,b))^2 \\
&= \frac{1}{2}\sum_{\alpha\in\Omega(\mu_{H\backslash\{e,h\}})}\mu_{H\backslash\{e,h\}}(\sigma)\sum_{a,b\in\Omega(\mu^\alpha_{e,f})}\mu^\sigma_{e,h}(a)\sum_{\beta\in\Omega(\mu^{\sigma,a}_{E\backslash H})\cap\Omega(\mu^{\sigma,b}_{E\backslash H})}\mu^{\alpha,a}_{E\backslash H}(\beta)(f(\alpha,\beta,a)-f(\alpha,\beta,b))^2 \\
&\leq \frac{1}{2}\sum_{\alpha\in\Omega(\mu_{H\backslash\{e,h\}})}\mu_{H\backslash\{e,h\}}(\sigma)\sum_{a,b\in\Omega(\mu^\alpha_{e,f})}\mu^\sigma_{e,h}(a)(g(\alpha,a)-g(\alpha,b))^2 \\
&\leq q^2\frac{1}{2}\sum_{\alpha\in\Omega(\mu_{H\backslash\{e,h\}})}\mu_{H\backslash\{e,h\}}(\sigma)\sum_{a,b\in\Omega(\mu^\alpha_{e,f})}\mu^\sigma_{e,h}(a)\mu^\sigma_{e,h}(b)(g(\alpha,a)-g(\alpha,b))^2 = q^2\nu[\mathbf{Var}_{\{e,h\}}[g]],
\end{aligned}
$$

where, in the last inequality, we use the fact that $\nu = \mu_H$ and $\mu^\sigma_{e,h}(b) \geq 1/q^2$. For the third case

47

where $e \in H$ and $h \notin H$, by a similar calculation, we have

$$\mu[\mathbf{Var}_{\{e,h\}}[f]]$$

$$= \frac{1}{2} \sum_{\sigma \in \Omega(\mu_{E \setminus \{e,h\}})} \mu_{E \setminus \{e,h\}}(\sigma) \sum_{a,b \in \Omega(\mu_{e,h}^{\sigma})} \mu_{e,h}^{\sigma}(a) \mu_{e,h}^{\sigma}(b) (f(\sigma,a) - f(\sigma,b))^2$$

$$= \frac{1}{2} \sum_{\sigma \in \Omega(\mu_{E \setminus \{e,h\}})} \mu_{E \setminus \{e,h\}}(\sigma) \sum_{a,b \in \Omega(\mu_e^{\sigma})} \mu_e^{\sigma}(a) \mu_e^{\sigma}(b) \sum_{x,y \in \Omega(\mu_h^{\sigma,a}) \cap \Omega(\mu_h^{\sigma,b})} \mu_h^{\sigma,a}(x) \mu_h^{\sigma,b}(y) (f(\sigma,a,x) - f(\sigma,b,y))^2$$

$$\leq \frac{1}{2} \sum_{\sigma \in \Omega(\mu_{E \setminus \{e,h\}})} \mu_{E \setminus \{e,h\}}(\sigma) \sum_{a,b \in \Omega(\mu_e^{\sigma})} \mu_e^{\sigma}(a) (f(\sigma,a,\cdot) - f(\sigma,b,\cdot))^2$$

$$= \frac{1}{2} \sum_{\alpha \in \Omega(\mu_{H \setminus \{e\}})} \mu_{H \setminus \{e\}}(\alpha) \sum_{a,b \in \Omega(\mu_e^{\alpha})} \mu_e^{\alpha}(a) \sum_{\beta \in \Omega(\mu_{E \setminus H \setminus \{h\}}^{\alpha,a}) \cap \Omega(\mu_{E \setminus H \setminus \{h\}}^{\alpha,b})} \mu_{E \setminus H \setminus \{h\}}^{\alpha,a}(\beta) (f(\alpha,\beta,a,\cdot) - f(\alpha,\beta,b,\cdot))^2$$

$$\leq \frac{1}{2} \sum_{\alpha \in \Omega(\mu_{H \setminus \{e\}})} \mu_{H \setminus \{e\}}(\alpha) \sum_{a,b \in \Omega(\mu_e^{\alpha})} \mu_e^{\alpha}(a) (g(\alpha,a) - g(\alpha,b))^2$$

$$\leq q \frac{1}{2} \sum_{\alpha \in \Omega(\mu_{H \setminus \{e\}})} \mu_{H \setminus \{e\}}(\alpha) \sum_{a,b \in \Omega(\mu_e^{\alpha})} \mu_e^{\alpha}(a) \mu_e^{\alpha}(b) (g(\alpha,a) - g(\alpha,b))^2 = q\nu[\mathbf{Var}_e[g]],$$

where, in the last inequality, we use the fact that $\nu = \mu_H$ and $\mu_e^{\alpha}(b) \geq 1/q$. $\qquad \square$

F   Proof of Lemma 26

Let $G = (V, E)$ be a graph and μ be a distribution over $[q]^E$ satisfying (23). Let $f : \Omega(\mu) \to \mathbb{R}$ be a function. First, we study some exchange property for the conditional expectation. For any two edges e and e', define $\mathrm{dist}(e, e')$ as the length of shortest path between e and e' in line graph. For the proofs in this section, we utilize the *conditional independence* in (23).

Proposition 51. *Let* $S, T \subseteq E$ *such that* $\mathrm{dist}(S, T) \geq 2$, *then it holds that*

$$\mu_{S \cup T}[f] = \mu_S \mu_T(f) = \mu_T \mu_S(f).$$

Proof. For a fixed $\eta \in \Omega(\mu)$, let $\rho = \eta_{E \setminus (S \cup T)}$. Note that ρ fixes the configuration both on ∂S and ∂T. By (23), conditional on ρ, configurations on S and T are independent. Then, by the definition of $\mu_S \mu_T[f]$ and above conditional independence property, one can verify

$$(\mu_S \mu_T(f))(\eta) = \sum_{\sigma \in \Omega(\mu_S^{\rho})} \mu_S^{\rho}(\sigma) \sum_{\tau \in \Omega(\mu_T^{\rho})} \mu_T^{\rho}(\tau) f(\sigma \uplus \tau \uplus \rho).$$

Note that σ, τ, ρ are configurations over three disjoint sets (which forms a partition of E), we use $\sigma \uplus \tau \uplus \rho$ to denote the merged configuration on E. Since RHS is symmetric with respect to S and T, we have $(\mu_S \mu_T(f))(\eta) = (\mu_T \mu_S(f))(\eta)$. Also, by definition and the conditional independence,

$$(\mu_{S \cup T}[f])(\eta) = \sum_{\sigma \in \Omega(\mu_S^{\rho})} \sum_{\tau \in \Omega(\mu_T^{\rho})} \mu_S^{\rho}(\sigma) \mu_T^{\rho}(\tau) f(\sigma \uplus \tau \uplus \rho). \qquad \square$$

Proposition 52. *If* $T \subseteq S \subseteq E$, *then it holds that*

$$\mu_S \mu_T(f) = \mu_T \mu_S(f) = \mu_S(f).$$

Proof. By the law of total expectation, we have $\mu_S \mu_T(f) = \mu_S(f)$. Also, note that $E \setminus S \subseteq E \setminus T$, by definition, it is easy to verify $\mu_T \mu_S(f) = \mu_S(f)$. $\qquad\square$

Proposition 53. *When $\partial S \cap T = \varnothing$, it holds that*

$$\mu_S \mu_T(f) = \mu_T \mu_S(f). \tag{67}$$

Proof. We let $A = S \setminus T$ and let $B = T \setminus S$. Since $\partial S \cap T = \varnothing$, it holds that $A \subseteq S$ and $\text{dist}(S, B) \geq 2$, which implies $\text{dist}(A, B) \geq 2$. According to Proposition 51 and Proposition 52, we have

$$\mu_S \mu_T(f) \overset{\text{Prop.51}}{=} \mu_S \mu_A \mu_B(f) \overset{\text{Prop.52}}{=} \mu_S \mu_B(f) \overset{\text{Prop.51}}{=} \mu_B \mu_S(f) \overset{\text{Prop.52}}{=} \mu_B \mu_A \mu_S(f) \overset{\text{Prop.51}}{=} \mu_T \mu_S(f). \quad \square$$

Finally, we are ready to prove Lemma 26. Let $S, T \subseteq E$ such that $\partial S \cap T = \varnothing$. We show that for any $f : \Omega(\mu) \to \mathbb{R}$,

$$\mu[\textbf{Var}_S [\mu_T[f]]] \leq \mu[\textbf{Var}_S [\mu_{S \cap T}[f]]].$$

For two functions g_1, g_2, we use $(g_1 - g_2)^2$ denote the function such that $(g_1 - g_2)^2(\sigma) = (g_1(\sigma) - g_2(\sigma))^2$. Note that for any function g, $\textbf{Var}_S [g] = \mu_S[(g - \mu_S[g])^2]$, where both LHS and RHS are functions and the equation here means equal two functions are the same. Using all above propositions, we have

$$
\begin{aligned}
\mu[\textbf{Var}_S [\mu_T[f]]] &= \mu \left[\mu_S[(\mu_T(f) - \mu_S \mu_T(f))^2] \right] \\
&= \mu \left[(\mu_T(f) - \mu_S \mu_T(f))^2 \right] \\
&\overset{(67)}{=} \mu \left[(\mu_T(f) - \mu_T \mu_S(f))^2 \right] \\
&= \mu \left[(\mu_T \mu_{S \cap T}(f) - \mu_T \mu_S \mu_{S \cap T}(f))^2 \right] \\
&\overset{(\star)}{\leq} \mu \left[(\mu_{S \cap T}(f) - \mu_S \mu_{S \cap T}(f))^2 \right] \\
&= \mu \left[\mu_S[(\mu_{S \cap T}(f) - \mu_S \mu_{S \cap T}(f))^2] \right] = \mu[\textbf{Var}_S [\mu_{S \cap T}[f]]],
\end{aligned}
$$

where $(\star)$ is an application of Jensen's inequality.