Sampling in Potts Model on Sparse Random Graphs*

Yitong Yin[†1] and Chihao Zhang[‡2]

1 **Nanjing University, Nanjing, China**
yinyt@nju.edu.cn

2 **Shanghai Jiao Tong University, Shanghai, China**
chihao.zhang@gmail.com

Abstract

We study the problem of sampling almost uniform proper q-colorings in sparse Erdős-Rényi random graphs $\mathcal{G}(n, d/n)$, a research initiated by Dyer, Flaxman, Frieze and Vigoda [2]. We obtain a fully polynomial time almost uniform sampler (FPAUS) for the problem provided $q > 3d + 4$, improving the current best bound $q > 5.5d$ [6].

Our sampling algorithm works for more generalized models and broader family of sparse graphs. It is an efficient sampler (in the same sense of FPAUS) for anti-ferromagnetic Potts model with activity $0 \le \beta < 1$ on $\mathcal{G}(n, d/n)$ provided $q > 3(1-\beta)d + 4$. We further identify a family of sparse graphs to which all these results can be extended. This family of graphs is characterized by the notion of contraction function, which is a new measure of the average degree in graphs.

1998 ACM Subject Classification F.2.2 Computations on Discrete Structures

Keywords and phrases Potts model, Sampling, Random Graph, Approximation Algorithm

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2016.47

1 Introduction

We study the problem of sampling almost uniform proper q-colorings in sparse Erdős-Rényi random graphs $\mathcal{G}(n, d/n)$. A classic sampling problem is to sample proper q-colorings of graphs with bounded *maximum degree* when $q \ge \alpha\Delta + \beta$, where Δ is the maximum degree. There is a substantial body of works on the problem [16, 1, 25, 3, 20, 14, 15, 13, 4, 10, 19]. The best positive result for this fundamental problem is the MCMC sampler for $q > \frac{11}{6}\Delta$ by Vigoda [25], and the best lower bound is due to Galanis, Štefankovič and Vigoda [9], which proved that the problem is intractable to solve when $q < \Delta$, even restricted to triangle-free Δ-regular graphs. The critical threshold $q = \Delta + 1$ is of great significance because it is the uniqueness threshold for the Δ-regular tree [18].

The studies of sampling proper q-colorings of graphs with bounded *average degree*, in particular the Erdős-Rényi random graph $\mathcal{G}(n, d/n)$ with constant d, was initiated in the seminal work of Dyer, Flaxman, Frieze and Vigoda [2], in which an algorithm was given to solve the problem with $q = \Theta(\log\log n / \log\log\log n)$ colors, substantially fewer than the maximum degree $\Theta(\log n / \log\log n)$ of the random graph. Several improvements have been

* This work was done in part while the authors were visiting the Simons Institute for the Theory of Computing.
† Yitong Yin is supported in part by NSFC grants 61272081 and 61321491.
‡ Chihao Zhang is supported in part by NSFC grant 61373029.

done since then. A significant step was made by Efthymiou and Spirakis [8] and independently by Mossel and Sly [21], in which the bound of q was improved to a constant $f(d)$ which is a large enough polynomial of d. Most recently, in a breakthrough by Efthymiou [6], an efficient algorithm was given to solve the problem when $q > 5.5d$, in linear of average degree d.

In all aforementioned results, the algorithms are FPAUSes (fully polynomial-time almost uniform samplers), meaning that for any $\epsilon > 0$ the algorithms terminates in time polynomial in n and $1/\epsilon$ and returns a random proper q-coloring according to a distribution within total variation distance ϵ from the uniform distribution over all proper q-colorings of the graph. For a much weaker goal where the total variation distance ϵ is fixed, an elegant combinatorial algorithm was given by Efthymiou [5, 7] to solve the problem for all $q > d$, approaching the uniqueness threshold.

In this paper, we consider FPAUS for proper q-colorings of $\mathcal{G}(n, d/n)$ with constant d. We give an algorithm which achieves an improved bound $q > 3d + O(1)$.

▶ **Theorem 1.** *For all sufficiently large constant d, all finite $q > 3d + 4$, there is an FPAUS for proper q-colorings of $G \sim \mathcal{G}(n, d/n)$ whp.*

The result is established in a more general context, namely the anti-ferromagnetic Potts model. In the q-state Potts model with activity β, given a graph $G = (V, E)$, a *configuration* $\sigma \in [q]^V$ assigns each vertex $v \in V$ one of the q colors from $[q]$, and is assigned with the weight

$$w(\sigma) = \prod_{uv \in E} \beta^{\mathbf{1}(\sigma(u)=\sigma(v))}.$$

The *Gibbs distribution* over all configurations $\sigma \in [q]^V$, denoted by $\mu = \mu_{q,\beta,G}$, is defined as $\mu(\sigma) = w(\sigma)/Z$ where the normalizing factor $Z = \sum_\sigma w(\sigma)$ is the *partition function*. When $0 \le \beta < 1$, the model is *anti-ferromagnetic*, meaning that adjacent vertices favor disagreeing colors. In particular, when $\beta = 0$ the Gibbs distribution is the uniform distribution over all proper q-colorings of G. In [19], it was discovered that sampling from Potts model is tractable for any q when $3(1-\beta)\Delta < \beta$, and the lower bound in [9] shows that it is intractable to sample in the anti-ferromagnetic Potts model on triangle-free Δ-regular graph for any even q when $q < (1-\beta)\Delta$.

We give the following sampling algorithm for anti-ferromagnetic Potts model on sparse random graphs.

▶ **Theorem 2.** *For all sufficiently large constant d, all $0 \le \beta < 1$ and $q > 3(1-\beta)d+4$, there is an algorithm such that for $G \sim \mathcal{G}(n, d/n)$ and any $\epsilon > 0$, the algorithm terminates in time polynomial in n and $1/\epsilon$ and returns a random q-coloring of G, according to a distribution within total variation distance ϵ from the Potts Gibbs distribution $\mu_{q,\beta,G}$ whp (with respect to the law of $\mathcal{G}(n, d/n)$).*

In particular, when $\beta = 0$, the above algorithm is an FPAUS for proper q-colorings of $G \sim \mathcal{G}(n, d/n)$ for $q > 3d + 4$. Theorem 1 is a special case of Theorem 2.

Our algorithm on graphs with bounded average degree asymptotically approaches the lower bound in [9] in terms of maximum degree.

In fact, the algorithm in Theorem 2 works for any family of graphs characterized by a particular *contraction function*. We introduce the notion of contraction function to generalize the *connective constant* [24, 23], a notion of average degree extensively studied in statistical physics. Therefore, the algorithm stated in Theorem 2 does not only work for Erdős-Rényi random graph but also for families of sparse graphs with a proper notion of bounded average degree. In particular, it holds for graphs with bounded maximum degree Δ when $q > 3(1-\beta)\Delta + 1$, which also greatly improves the existing upper bounds for anti-ferromagnetic Potts model on graphs with bounded maximum degree [10, 19]. The definition of contraction function and the full statement of the main result with it are quite technical. We defer them to Section 2.

1.1 Techniques

In most of the previous works [2, 8, 21, 6], the sampling algorithms were based on block Glauber dynamics. For proper q-colorings, if the degree of a vertex is much higher than q, then the standard Glauber dynamics will have torpid mixing around that vertex since the color of that vertex will be frozen for most of the time. In previous works this was overcome by using block dynamics, such that within a block the high-degree vertices are hidden in the block's "core", which is separated from the block's boundary by an intermediate "buffer" of low-degree vertices. It is not hard to imagine that the construction of such blocks can be quite complicated and the efficient construction of blocks crucially relies on the sparsity of Erdős-Rényi random graph $\mathcal{G}(n, d/n)$.

In contrast, we use the correlation decay technique. This approach was introduced to multi-spin models (e.g. colorings) in the seminal work of Gamarnik and Katz [10], in which they gave an FPAUS for proper q-colorings when $q > 2.844\Delta$ where Δ is the maximum degree. This was later improved to $q > 2.581\Delta$ in [19], which remains to be the best bound achieved by correlation-decay-based algorithms for proper q-colorings.

Our algorithm heavily relies on the computation tree recursion introduced in [10]. The basic idea is simple: sampling with the estimations of marginal probabilities, which are computed approximately by a proper truncation of the the computation tree recursion. With correlation decay, the approximation is accurate enough so the algorithm is an FPAUS. A complication here is that the degrees of vertices are unbounded. We overcome this by introducing a computation tree in terms of blocks and establish the decay of correlation between blocks.

The blocks in our algorithm can be constructed straightforwardly: they are just clusters of high-degree vertices. Due to the simple and generic construction of blocks, our algorithm may work for general families of graphs, and can be applied as a generic method for graphs with a few high-degree vertices.

The idea of block correlation decay was introduced in our previous work [26] to establish the correlation decay for proper q-colorings of $\mathcal{G}(n, d/n)$ for $q > 2d + O(1)$, by a block modification to another recursion of Gamarnik, Katz and Misra [11]. This recursion is suitable for proving "correlation decay only" result. A drawback of the current approaches based on correlation decay is that we do not know how to use this approach to get an algorithm achieving a bound which is close to $q > 2\Delta + O(1)$, even on graphs with maximum degree Δ. Sampling proper q-coloring in $\mathcal{G}(n, d/n)$ for $q > 2d$ or smaller q may require new understandings of correlation decay in multi-spin systems, or may have to use other techniques such as Glauber dynamics.

2 Preliminary and statement of the main result

Let $G = (V, E)$ be an undirected graph. For any subset $S \subseteq V$ of vertices, let $G[S]$ denote the subgraph of G induced by S, and let $\partial B = \{u \in V \setminus B \mid \exists w \in B, (u, w) \in E\}$ denote the vertex boundary of B. Given a vertex v in G, let $\mathrm{dist}_G(v, S)$ denote the minimum distance from v to any vertex $u \in S$ in G. In case that $S = \{u\}$ is a singleton, we write $\mathrm{dist}_G(v, u)$ instead of $\mathrm{dist}_G(v, S)$.

Potts model

The anti-ferromagnetic Potts model is parameterized by an integer $q \geq 2$ and an *activity* parameter $0 \leq \beta < 1$. Each element of $[q]$ represents a *color* or a *state*. Let $G = (V, E)$ be a graph. A *configuration* $\sigma \in [q]^\Lambda$ on a subset $\Lambda \subseteq V$ of vertices assigns each vertex v in Λ one of the q colors in $[q]$. In the Potts model on graph G, each configuration $\sigma \in [q]^V$ is assigned a weight

$$w_G(\sigma) = \beta^{\#\mathsf{mon}(\sigma)},$$

where $\#\mathsf{mon}(\sigma) = |\{(u, v) \in E \mid \sigma(u) = \sigma(v)\}|$ gives the number of monochromatic (undirected) edges in the configuration σ.

The analysis of correlation decay introduces Potts model with boundary conditions. More formally, we consider an instance of Potts model as a tuple $\Omega = (G, \Lambda, \sigma)$ where $G = (V, E)$ is an undirected graph, $\Lambda \subseteq V$ is a subset of vertices in G and $\sigma \in [q]^\Lambda$ is a configuration on Λ. Given such an instance $\Omega = (G, \Lambda, \sigma)$, the weight function w_Ω assigns each configuration $\pi \in [q]^V$ the weight $w_\Omega(\pi) = w_G(\pi)$ if π agrees with σ over all vertices in Λ, and $w_\Omega(\pi) = 0$ if otherwise. An instance Ω is *feasible* if there exists a configuration on V with positive weight. This gives rise to a natural probability distribution $\mu = \mu_{q,\beta,G}$, called *Gibbs distribution*, over all configurations $\pi \in [q]^V$ for a feasible Potts instance:

$$\mu(\pi) = \mathbf{Pr}_\Omega[c(V) = \pi] = \frac{w_\Omega(\pi)}{Z(\Omega)},$$

where $Z(\Omega) = \sum_{\sigma \in [q]^V} w_\Omega(\sigma)$ is the *partition function.* For a vertex $v \in V$ and any color $x \in [q]$, we use $\mathbf{Pr}_\Omega[c(v) = x]$ to denote the marginal probability that v is assigned color x by a configuration sampled from the Gibbs distribution. Similarly, for a set $S \subseteq V$ and $\pi \in [q]^S$, we use $\mathbf{Pr}_\Omega[c(S) = \pi]$ to denote the marginal probability that S is assigned configuration π by a configuration sampled from the Gibbs distribution.

Block and sparsity

Fix any $q \geq 2$ and $0 \leq \beta < 1$. Let $\Omega = (G, \Lambda, \sigma)$ be an instance of q-state Potts model with activity β and v a vertex in G. We call v a *low-degree* vertex if $\deg_G(u) < \frac{q-1}{1-\beta} - 2$, and otherwise we call it a *high-degree* vertex.

▶ **Definition 3** (permissive block)**.** Let $\Omega = (G, \Lambda, \sigma)$ be a Potts instance where $G = (V, E)$. A vertex set $B \subseteq V \setminus \Lambda$ is a *permissive block* in Ω if every boundary vertex $u \in \partial B \setminus \Lambda$ is a low-degree vertex. For any subset of vertices $S \subseteq V \setminus \Lambda$, we denote $B(S) = B_\Omega(S)$ the *minimal permissive block* containing S. We write $B(v) = B(S)$ if $S = \{v\}$ is a singleton.

▶ **Definition 4.** A family $\mathcal{G}$ of finite graphs is *locally sparse* if there exists a constant $C > 0$ such that for every $G = (V, E)$ in the family and every path P in G of length ℓ we have $|B(P)| \leq C(\ell + \log |V|)$.

SAW tree

Given a graph $G = (V, E)$ and a vertex $v \in V$, a rooted tree T can be naturally constructed from all self-avoiding walks starting from v in G as follows: Each vertex in T corresponds to a self-avoiding walk (simple path in G) $P = (v, v_1, v_2, \ldots, v_k)$ starting from v, whose children correspond to all self-avoiding walks $(v, v_1, v_2, \ldots, v_k, v_{k+1})$ in G extending P, and the root of T corresponds to the trivial walk (v). The resulting tree, denoted by $T_{\text{SAW}}(G, v)$, is called the *self-avoiding walk* (SAW) tree constructed from vertex v in graph G.

From this construction, every vertex in $T_{\text{SAW}}(G, v)$ can be naturally identified with the vertex in V (many-to-one) at which the corresponding self-avoiding walk ends.

Contraction function

Given a vertex v in a locally finite graph $G = (V, E)$, let $\mathsf{SAW}(v, \ell)$ denote the set of self-avoiding walks in G of length ℓ starting at v. The following notion of connective constant of families of finite graphs is introduced in [24].

▶ **Definition 5** (connective constant [24, 23])**.** Let $\mathcal{G}$ be a family of finite graphs. The *connective constant* of $\mathcal{G}$ is bounded by Δ if there exists a positive constant $C > 0$ such that for any graph $G = (V, E)$ in $\mathcal{G}$ and any vertex v in G, we have $|\mathsf{SAW}(v, \ell)| \leq n^C \Delta^\ell$ where $n = |V|$ for all $\ell \geq 1$.

Let $\delta : \mathbb{N} \to \mathbb{R}^+$ be a function. Given a vertex v in a locally finite graph $G = (V, E)$, let

$$\mathcal{E}_\delta(v, \ell) := \sum_{\substack{(v, v_i, \ldots, v_\ell) \\ \in \mathsf{SAW}(v, \ell)}} \prod_{i=1}^{\ell} \delta(\deg(v_i)). \tag{1}$$

▶ **Definition 6** (contraction function)**.** Let $\mathcal{G}$ be a family of finite graphs. The $\delta : \mathbb{N} \to \mathbb{R}^+$ is a *contraction function* for $\mathcal{G}$ if there exist positive constants $C > 0, \gamma < 1$ such that for any graph $G = (V, E)$ in $\mathcal{G}$ and any vertex v in G, we have $\mathcal{E}_\delta(v, \ell) < n^C \gamma^\ell$ where $n = |V|$ for all $\ell \geq 1$.

It is easy to see that graph families $\mathcal{G}$ with constant contraction function $\delta(d) = \frac{1}{\Delta}$ are precisely the families $\mathcal{G}$ of connective constant bounded strictly by Δ.

Statement of the main result

Now we are ready to state our main technical result.

▶ **Theorem 7** (Main theorem)**.** *Let $q \geq 3$ be an integer and $0 \leq \beta < 1$. Let $\mathcal{G}$ be a family of finite graphs that satisfies the followings:*

- *the following $\delta(\cdot)$ is a contraction function for $\mathcal{G}$:*

$$\delta(d) = \begin{cases} \frac{2(1-\beta)}{q-1-(1-\beta)d} & \text{if } d \leq \frac{q-1}{1-\beta} - 2, \\ 1 & \text{otherwise;} \end{cases} \tag{2}$$

- *$\mathcal{G}$ is locally sparse;*
- *(proper q-coloring) if $\beta = 0$, then $\mathcal{G}$ also needs to be q-colorable.*

Then there is an FPTAS for computing the partition function $Z(\Omega)$ for every $\Omega = (G, \Lambda, \sigma)$ with $G \in \mathcal{G}$. Consequently, there is an algorithm such that for all $G = (V, E) \in \mathcal{G}$, all $\epsilon > 0$, the algorithm terminates in time polynomial in $n = |V|$ and $1/\epsilon$, and returns a random $\sigma \in [q]^V$ according a distribution within total variation distance ϵ from the Potts Gibbs distribution $\mu_{q,\beta,G}$.

3 The computation tree for blocks

In this section, we introduce recursions to compute the marginal probabilities on a vertex and on a permissive block respectively.

When $\beta = 0$, the model becomes proper q-coloring, and the feasibility of a configuration becomes an issue.

Let $\Omega = (G, \Lambda, \sigma)$, where $G = (V, E)$ be a feasible instance of proper q-coloring. Recall that an instance $\Omega = (G, \Lambda, \sigma)$ is feasible if there exists a proper q-coloring consistent with σ. For a subset of vertices $S \subseteq V \setminus \Lambda$, a q-coloring $\pi \in [q]^S$ is (globally) *feasible* if it can be extended to a proper q-coloring of G. A q-coloring $\pi \in [q]^S$ is *locally feasible*, if $\sigma \cup \pi$ is a proper q-coloring in the subgraph $G[\Lambda \cup S]$ induced by $\Lambda \cup S$.

▶ **Proposition 8.** *Let $\Omega = (G, \Lambda, \sigma)$ where $G = (V, E)$ be a feasible instance of proper q-coloring, $v \in V \setminus \Lambda$ be a vertex and $\pi \in [q]^{B(v)}$ be a locally feasible configuration. Then π is also feasible.*

Proof. Denote $B = B(v)$. Fix a configuration $\eta \in [q]^V$ such that $w_\Omega(\eta) > 0$, this is possible since Ω is feasible. We denote by η' the restriction of η to $V \setminus ((B \cup \partial B) \setminus \Lambda)$, i.e., the set of vertices that are either in Λ, or not in $B \cup \partial B$.

Consider the configuration $\eta = \pi \cup \eta' \in [q]^{V \setminus (\partial B \setminus \Lambda)}$, it can be extended to a configuration $\rho \in [q]^V$ with $w_\Omega(\rho) > 0$ in a greedy fashion, since every vertex in $\partial B \setminus \Lambda$ is of low-degree. Thus ρ witness that π is feasible. ◀

With this proposition, we do not distinguish between local feasibility and feasibility of configurations on permissive blocks. For a permissive block B, we use $\mathcal{F}(B)$ to denote the set of feasible configuration. Note that when $\beta > 0$, the set $\mathcal{F}(B)$ is simply $[q]^B$.

3.1 The recursion

Let $\Omega = (G, \Lambda, \sigma)$ where $G = (V, E)$ be an instance of Potts model and $v \in V \setminus \Lambda$ be a vertex. Let $B = B(v)$ be the minimal permissive block containing v. Let $\delta B = \{u_i v_i \mid i \in [m]\}$ be an enumeration of boundary edges of B where $v_i \notin B$ for every $i \in [m]$. In this notation, more than one u_i or v_i may refer to the same vertex. We denote $E(B) := \{uv \in E \mid u, v \in B\}$ the edges in B. We use $\bar{B}$ to denote the inner boundary of B, i.e., $\bar{B} = \{u \in B \mid uv \in E \text{ and } v \notin B\}$.

Recall that we use $\mathcal{F}(B)$ to denote the set of feasible configurations on a permissive block B, it is easy to see that, for every $x \in [q]$,

$$\mathbf{Pr}_\Omega\left[c(v) = x\right] = \sum_{\substack{\pi \in \mathcal{F}(B): \\ \pi(v) = x}} \mathbf{Pr}_\Omega\left[c(B) = \pi\right].$$

This identity relates the marginal probability on a vertex to marginal probabilities on a block. We now define notations for some sub-instances and give a block-to-vertices identity.

Let $\pi \in \mathcal{F}(B)$ be a configuration on a permissive block B. For every $i \in [m]$, denote $\pi_i = \pi(u_i)$. Let $G_B = (V_B, E_B)$ denote the graph obtained from G by removing $B \setminus \bar{B}$ and edges in $E(B)$, i.e., $V' = (V \setminus B) \cup \bar{B}$, $E' = E \setminus E(B)$. Let $\Omega_B = (G_B, \Lambda, \sigma)$. For every $i = 1, 2, \ldots, m+1$, define $\Omega_i^\pi = (G_i^\pi, \Lambda_i^\pi, \sigma_i^\pi)$ as the instance obtained from Ω_B by fixing u_j to color π_j for every $j \in [i-1]$ and by removing edges $u_j v_j$ for every $j = i, i+1, \ldots, m$.

▶ **Lemma 9.** *Assuming above notations, it holds that*

$$\mathbf{Pr}_{\Omega}\left[c(B)=\pi\right]=\frac{w_{G[B]}(\pi)\cdot\prod_{i=1}^{m}\left(1-(1-\beta)\mathbf{Pr}_{\Omega_i^{\pi}}\left[c(v_i)=\pi_i\right]\right)}{\sum_{\rho\in\mathcal{F}(B)}w_{G[B]}(\rho)\cdot\prod_{i=1}^{m}\left(1-(1-\beta)\mathbf{Pr}_{\Omega_i^{\rho}}\left[c(v_i)=\rho_i\right]\right)}. \tag{3}$$

Proof.

$$\begin{aligned}\mathbf{Pr}_{\Omega}\left[c(B)=\pi\right]&=\frac{w_{G[B]}(\pi)\cdot Z(\Omega_{m+1}^{\pi})}{\sum_{\rho\in\mathcal{F}(B)}w_{G[B]}(\rho)\cdot Z(\Omega_{m+1}^{\rho})}=\frac{w_{G[B]}(\pi)\cdot\frac{Z(\Omega_{m+1}^{\pi})}{Z(\Omega_1^{\pi})}}{\sum_{\rho\in\mathcal{F}(B)}w_{G[B]}(\rho)\cdot\frac{Z(\Omega_{m+1}^{\rho})}{Z(\Omega_1^{\rho})}}\\&=\frac{w_{G[B]}(\pi)\cdot\prod_{i=1}^{m}\frac{Z(\Omega_{i+1}^{\pi})}{Z(\Omega_i^{\pi})}}{\sum_{\rho\in\mathcal{F}(B)}w_{G[B]}(\rho)\cdot\prod_{i=1}^{m}\frac{Z(\Omega_{i+1}^{\rho})}{Z(\Omega_i^{\rho})}}.\end{aligned}$$

Since for every $\rho\in\mathcal{F}(B)$ and $i\in[d]$,

$$Z(\Omega_{i+1}^{\rho})=\sum_{y\in[q]}Z\left(\Omega_i^{\rho}\mid c(v_i)=y\right)\cdot\beta^{\mathbf{1}(y=\rho(u_i))}$$

where $Z\left(\Omega_i^{\rho}\mid c(v_i)=y\right)$ stands for the sum of the weights of all feasible configurations σ on Ω_i^{ρ} satisfying $\sigma(v_i)=y$ and $\mathbf{1}(\cdot)$ is the indicator function. With this identity, we can further write

$$\begin{aligned}\mathbf{Pr}_{\Omega}\left[c(B)=\pi\right]&=\frac{w_{G[B]}(\pi)\cdot\prod_{i=1}^{m}\frac{\sum_{y\in[q]}Z(\Omega_i^{\pi}\mid c(v_i)=y)\cdot\beta^{\mathbf{1}(y=\pi(u_i))}}{Z(\Omega_i^{\pi})}}{\sum_{\rho\in\mathcal{F}(B)}w_{G[B]}(\rho)\cdot\prod_{i=1}^{m}\frac{\sum_{y\in[q]}Z(\Omega_i^{\rho}\mid c(v_i)=y)\cdot\beta^{\mathbf{1}(y=\rho(u_i))}}{Z(\Omega_i^{\rho})}}\\&=\frac{w_{G[B]}(\pi)\cdot\prod_{i=1}^{m}\left(1-(1-\beta)\mathbf{Pr}_{\Omega_i^{\pi}}\left[c(v_i)=\pi_i\right]\right)}{\sum_{\rho\in\mathcal{F}(B)}w_{G[B]}(\rho)\cdot\prod_{i=1}^{m}\left(1-(1-\beta)\mathbf{Pr}_{\Omega_i^{\rho}}\left[c(v_i)=\rho_i\right]\right)}.\end{aligned}$$ ◀

This identity expresses the marginal probability on a permissive block as the function of marginal probabilities on its incident vertices, with modified instances. We now analyze the derivatives of this function, which is important in the analysis of correlation decay.

▶ **Lemma 10.** *Let* $\mathbf{p}=(p_{i,\rho})_{i\in[m],\rho\in\mathcal{F}(B)}$, $\hat{\mathbf{p}}=(\hat{p}_{i,\rho})_{i\in[m],\rho\in\mathcal{F}(B)}$ *be two tuples of variables and*

$$f(\mathbf{p}):=\frac{w_{G[B]}(\pi)\prod_{i=1}^{m}\left(1-(1-\beta)p_{i,\pi}\right)}{\sum_{\rho\in\mathcal{F}(B)}w_{G[B]}(\rho)\prod_{i=1}^{m}\left(1-(1-\beta)p_{i,\rho}\right)}.$$

Assume for every $i\in[m],\rho\in\mathcal{F}(B(v))$, $p_{i,\rho},\hat{p}_{i,\rho}\le\frac{1-\beta}{q-(1-\beta)d_i}$, *then*

$$\left|\log f(\mathbf{p})-\log f(\hat{\mathbf{p}})\right|\le\sum_{i\in[d]}\frac{2(1-\beta)}{q-(1-\beta)d_i-1}\cdot\max_{\rho\in\mathcal{F}(B(v))}\left|\log p_{i,\rho}-\log\hat{p}_{i,\rho}\right|.$$

Proof. For every $i\in[m]$, we have

$$\frac{\partial f}{\partial p_{i,\pi}}=-(1-\beta)f(1-f)\cdot\frac{1}{1-(1-\beta)p_{i,\pi}}.$$

For every $i\in[m]$ and $\rho\neq\pi$, we have

$$\frac{\partial f}{\partial p_{i,\rho}}=(1-\beta)f\cdot\frac{w_{G[B]}(\rho)\prod_{i=1}^{m}(1-(1-\beta)p_{i,\rho})}{\sum_{\sigma\in\mathcal{F}(B)}w_{G[B]}(\sigma)\prod_{i=1}^{m}(1-(1-\beta)p_{i,\sigma})}\cdot\frac{1}{1-(1-\beta)p_{i,\rho}}.$$

Thus,

$$\sum_{\substack{\rho\in\mathcal{F}(B)\\ \rho\neq\pi}} \frac{\partial f}{\partial p_{i,\rho}} \leq (1-\beta)f(1-f)\cdot \max_{\substack{\rho\in\mathcal{F}(B)\\ \rho\neq\pi}} \frac{1}{1-(1-\beta)p_{i,\rho}}.$$

Let $\Phi = \frac{1}{x}$, by mean value theorem, for some $\tilde{\mathbf{p}} = (\tilde{p}_{i,\rho})_{i\in[m],\rho\in\mathcal{F}(B(v))}$ where each $\tilde{p}_{i,\rho} \leq \frac{1-\beta}{q-(1-\beta)d_i}$, we have

$$\begin{aligned}
&|\log f(\mathbf{p}) - \log f(\hat{\mathbf{p}})| \\
&= \sum_{i\in[m]}\sum_{\rho\in\mathcal{F}(B)} \left(\frac{\Phi(f)}{\Phi(p_{i,\rho})}\left|\frac{\partial f}{\partial p_{i,\rho}}\right|\right)\Bigg|_{\mathbf{p}=\tilde{\mathbf{p}}} \cdot |\log p_{i,\rho} - \log \hat{p}_{i,\rho}| \\
&\leq \sum_{i\in[m]} \left(\frac{\Phi(f)}{\Phi(p_{i,\pi})}\left|\frac{\partial f}{\partial p_{i,\pi}}\right| + \sum_{\substack{\rho\in\mathcal{F}(B)\\ \rho\neq\pi}} \frac{\Phi(f)}{\Phi(p_{i,\rho})}\left|\frac{\partial f}{\partial p_{i,\rho}}\right|\right)\Bigg|_{\mathbf{p}=\tilde{\mathbf{p}}} \cdot \max_{\rho\in\mathcal{F}(B(v))} |\log p_{i,\rho} - \log \hat{p}_{i,\rho}| \\
&\leq \sum_{i\in[m]} \left((1-\beta)\left(\frac{p_{i,\pi}}{1-(1-\beta)p_{i,\pi}} + \max_{\substack{\rho\in\mathcal{F}(B)\\ \rho\neq\pi}} \frac{p_{i,\rho}}{1-(1-\beta)p_{i,\rho}}\right)\right)\Bigg|_{\mathbf{p}=\tilde{\mathbf{p}}} \\
&\qquad \cdot \max_{\rho\in\mathcal{F}(B(v))} |\log p_{i,\rho} - \log \hat{p}_{i,\rho}| \\
&\leq \sum_{i\in[m]} \frac{2(1-\beta)}{q-(1-\beta)d_i-(1-\beta)} \cdot \max_{\rho\in\mathcal{F}(B(v))} |\log p_{i,\rho} - \log \hat{p}_{i,\rho}|.
\end{aligned}$$

◀

3.2 Bounds for marginals

The following lemma gives an upper bound for the probability $\mathbf{Pr}_\Omega\left[c(v)=x\right]$.

▶ **Lemma 11.** *Assume $q > (1-\beta)d$. For every color $x\in[q]$, it holds that*

$$\mathbf{Pr}_\Omega\left[c(v)=x\right] \leq \frac{1}{q-(1-\beta)d},$$

where d is the degree of v in G.

Proof. Assume $x=1$. For every $i\in[q]$, let x_i denote the number of neighbors of v that are of color i. Then $p_{v,1} \leq \max \frac{\beta^{x_1}}{\sum_{i\in[q]}\beta^{x_i}}$ subject to the constraints that all x_i are nonnegative integers and $\sum_{i=1}^{q} x_i = d$. Since $\beta\leq 1$, we can assume $x_1=0$, thus $p_{v,1}\leq \max\frac{1}{1+\sum_{i=2}^{q}\beta^{x_i}}$. We now distinguish between two cases:

1. (If $d\geq q-1$) In this case, let $\lambda = 1-\beta$, then

$$\frac{1}{1+\sum_{i=2}^{q}\beta^{x_i}} \leq \frac{1}{1+(q-1)(1-\lambda)^{\frac{d}{q-1}}} \overset{\heartsuit}{\leq} \frac{1}{1+(q-1)\left(1-\frac{\lambda d}{q-1}\right)} = \frac{1}{q-(1-\beta)d},$$

   where $\heartsuit$ is due to the fact that the inequality $(1-a)^b \geq 1-ab$ holds when $0\leq a\leq 1$ and $b\geq 1$.
2. (If $d < q-1$) In this case, due to the integral constraint of x_i's, the term $\sum_{i=2}^{q}\beta^{x_i}$ minimizes when d of x_i's are set to one and remaining x_i's are set to zero. Therefore, we have

$$\frac{1}{1+\sum_{i=2}^{q}\beta^{x_i}} \leq \frac{1}{1+d\beta+(q-1-d)} = \frac{1}{q-(1-\beta)d},$$

◀

Algorithm 1: $\texttt{marg}(\Omega, v, x, \ell)$

If v is fixed to be color y, then return 1 if $x = y$ and return 0 if $x \neq y$;
If $\ell < 0$ return $1/q$;
Compute $B(v)$;
For every $\rho \in \mathcal{F}(B(v))$, let $\hat{p}_\rho \leftarrow \texttt{marg-block}(\Omega, B(v), \rho, \ell)$;
Return $\min\left\{ \sum_{\substack{\pi \in \mathcal{F}(B(v)) \\ s.t.\ \pi(v)=x}} \hat{p}_\pi, \frac{1}{\max\{1, q-(1-\beta)\deg_G(v)\}} \right\}$

Algorithm 2: $\texttt{marg-block}(\Omega, B(v), \pi, \ell)$

Compute P_i for every $i \in [m]$;
$\hat{p}_{i,\rho} \leftarrow \texttt{marg}(\Omega_i^\rho, v_i, \rho_i, \ell - |P_i|)$ for every $i \in [m]$ and $\rho \in \mathcal{F}(B)$;
Return $\frac{w_{G[B]}(\pi) \prod_{i \in [m]} (1-(1-\beta)\hat{p}_{i,\pi})}{\sum_{\rho \in \mathcal{F}(B)} w_{G[B]}(\rho) \prod_{i \in [m]} (1-(1-\beta)\hat{p}_{i,\rho})}$;

The recursion (3) holds for arbitrary set of vertices B (not necessary a permissive block), thus if one takes B as a single vertex, it implies the following simple lower bound for marginal probabilities on a vertex.

▶ **Lemma 12.** *For every feasible $x \in [q]$, it holds that*

$$\mathbf{Pr}_\Omega[c(v) = x] \geq \frac{\beta^d}{q},$$

where d is the degree of v in G.

3.3 The algorithm

We now implement the recursions introduced in previous sections to estimate marginals. There is a slight difference between the case of $\beta > 0$ and the case of $\beta = 0$. If $\beta = 0$, our algorithm may encounter an infeasible instance and we need to check the feasibility in advance.

The $\beta > 0$ case

We define two procedures $\texttt{marg}(\Omega, v, x, \ell)$ and $\texttt{marg-block}(\Omega, B(v), \pi, \ell)$ calling each other to estimate vertex and block marginal respectively. We assume $\Omega = (G, \Lambda, \sigma)$ with $G = (V, E)$ is an instance of Potts model, $v \in V \setminus \Lambda$ is a vertex, $x \in [q]$ is a color and ℓ is an integer. Recall that for a permissive block $B(v)$, we use $\mathcal{F}(B)$ to denote the set of feasible configurations over $B(v)$.

To describe the algorithm for estimating the block marginals, we need to introduce some notations. Let $B = B(v)$, and we enumerate the boundary edges in δB by $e_i = u_i v_i$ for $i = 1, 2, \ldots, m$, where $v_i \notin B$. With this notation more than one u_i or v_i may refer to the same vertex, which is fine. For every $i \in [m]$ and $\rho \in \mathcal{F}(B)$, define Ω_B and Ω_i^ρ as in Lemma 9.

Let $P_i = (v, w_1, w_2, \ldots, w_k, v_i)$ be a self-avoiding walk from v to v_i such that all intermediate vertices w_i are in $B(v)$. Since $B(v)$ is a minimal permissive block, such walk always exists, and let P_i be an arbitrary one of them if there are multiple ones.

Algorithm 3: $\texttt{marg}(\Omega, v, x, \ell)$

1. If v is fixed to be color y, then return 1 if $x = y$ and return 0 if $x \neq y$;
2. Compute $B(v)$;
3. If $\ell < 0$, then return $1/q$ if there is a feasible $\pi \in \mathcal{F}(B(v))$ such that $\pi(v) = x$ and return 0 if no such π exists;
4. For every $\rho \in \mathcal{F}(B(v))$, let $\hat{p}_\rho \leftarrow \texttt{marg-block}(\Omega, B(v), \rho, \ell)$;
5. Return $\min\left\{\sum_{\substack{\pi \in \mathcal{F}(B(v)) \\ s.t.\ \pi(v)=x}} \hat{p}_\pi, \frac{1}{\max\{1, q-(1-\beta)\deg_G(v)\}}\right\}$

The $\beta = 0$ case

We slightly modify our procedure to deal with infeasible instance. Let $\Omega = (G, \Lambda, \sigma)$ be an instance of Potts model with $q \geq 3$ and activity $\beta = 0$ where $G = (V, E)$, $v \in V \setminus \Lambda$ be a vertex, $x \in [q]$ be a color and ℓ be an integer. We define

The only difference of this version of $\texttt{marg}$ is at step 3, where we check whether the color x is locally feasible. We return $1/q$ if so and return 0 otherwise.

4 Correlation decay

In this section, we show that the algorithms introduced in Section 3 to estimate marginals are accurate, if the input instance satisfies the conditions specified in the statement of Theorem 7.

▶ **Lemma 13.** *Let $q \geq 3$ be an integer and $0 \leq \beta < 1$ be a real. Let $\mathcal{G}$ be a family of finite graphs satisfying the conditions of Theorem 7.*

There exists an algorithm such that for every feasible instance $\Omega = (G, \Lambda, \sigma)$ of Potts model where $G = (V, E) \in \mathcal{G}$ with $|V| = n$, $\Lambda \subseteq V$ and $\sigma \in [q]^\Lambda$, for every vertex $v \in V$ and every color $x \in [q]$, it can compute an estimation $\hat{p}$ of $\mathbf{Pr}_\Omega[c(v) = x]$ in time polynomial in n, such that

$$1 - O\left(\frac{1}{n^3}\right) \leq \frac{\hat{p}}{\mathbf{Pr}_\Omega[c(v) = x]} \leq 1 + O\left(\frac{1}{n^3}\right).$$

To prove Lemma 13, we introduce the notion of error function to relate contraction function and the accurate of our estimation algorithm.

▶ **Definition 14.** Given an instance $\Omega = (G, \Lambda, \sigma)$ of Potts model with $q \geq 3$ and activity $0 \leq \beta < 1$ where $G = (V, E)$ with $|V| = n$. Let $v \in V \setminus \Lambda$ be a vertex, $T = T_{\mathrm{SAW}}(G[V], v)$ be the self-avoiding walk tree rooted at v in G and S be a set of vertices in T. Assume v has m children $v_1, v_2, \ldots, v_m$ in T, let T_i denote the subtree of T rooted at v_i. We recursively define the error function:

- Case $\beta > 0$:

$$\mathcal{E}_{T,S} := \begin{cases} \sum_{i=1}^m \delta(\deg_G(v_i)) \cdot \mathcal{E}_{T_i,S} & \text{if } v \notin S \cup \Lambda, \\ q + n \log \frac{1}{\beta} & \text{if } v \in S, \\ 0 & \text{if } v \in \Lambda. \end{cases}$$

- Case $\beta = 0$:

$$\mathcal{E}_{T,S} := \begin{cases} \sum_{i=1}^m \delta(\deg_G(v_i)) \cdot \mathcal{E}_{T_i,S} & \text{if } v \notin S \cup \Lambda, \\ n \log q & \text{if } v \in S, \\ 0 & \text{if } v \in \Lambda. \end{cases}$$

In the above definition, the set S specifies the boundary of our recursively defined error function $\mathcal{E}_{T,S}$. The error function $\mathcal{E}_{T,S}$ will be used as an upper bound for the error in our estimation algorithm. If the function $\delta(\cdot)$ is a contraction for a family $\mathcal{G}$, then for every graph $G = (V, E) \in \mathcal{G}$ and vertex $v \in V$, as we shall show in the next lemma, there exists a set of low-degree vertices in $T_{\text{SAW}}(G, v)$ at certain depth. This set of vertices, as it will become clear later, serves as the boundary S in our computation tree.

▶ **Lemma 15.** *Let $\mathcal{G}$ be a family of finite graphs for which $\delta(\cdot)$ is a contraction function. Then for some constants $\theta > 1$ and $C > 0$, for every $G = (V, E) \in \mathcal{G}$ with $|V| = n$, every $v \in V$ and every $L \geq C \log n$, there exists a low-degree S in $T = T_{\text{SAW}}(G, v)$ such that for every $u \in S$, $L < \text{dist}_T(u, v) \leq \theta L$ and every self-avoiding walk in T from v of length θL intersects S.*

Proof. Let $G = (V, E) \in \mathcal{G}$ be a graph. It follows from the definition of contraction function that for some constant $C > 0$, for every $\ell \geq C \log n$, $\mathcal{E}_\delta(v, \ell) < \alpha^\ell$ for some constant $0 < \alpha < 1$.

It is sufficient to show that, for some constant integer $\theta > 0$ it holds that for every $v \in V$, every $L \geq C \log n$, every $P = (v, v_1, \ldots, v_{\theta L}) \in \mathsf{SAW}(v, \theta L)$, there exists a low-degree vertex v_j among $\{v_{L+1}, v_{\theta L}, \ldots, v_{\theta L}\}$.

Let $\theta = \max\left\{\lceil \log_{1/\alpha}\left(\frac{q-1}{2(1-\beta)}\right)\rceil, 2\right\}$. Assume for the contradiction that every vertex in $\{v_{L+1}, v_{L+2}, \ldots, v_{\theta L}\}$ has high-degree. Since $\theta L > L \geq C \log n$, we have $\prod_{i=1}^{\theta L} \delta(\deg(v_i)) \leq \alpha^{\theta L}$.

On the other hand, since $\delta(d) \geq \delta(0) = \frac{2(1-\beta)}{q-1}$, we have $\prod_{i=1}^{\theta L} \delta(\deg(v_i)) \geq \left(\frac{2(1-\beta)}{q-1}\right)^L$. This is a contradiction for our choice of θ. ◀

We now define the error of our estimation.

▶ **Definition 16.** Let $\Omega = (G, \Lambda, \sigma)$ with $G = (V, E)$ be an instance of Potts model, $v \in V \setminus \Lambda$ be a vertex, $x \in [q]$ be a color and $\ell \in \mathbb{Z}$ be an integer. Let $\hat{p}_{\Omega,v,x,\ell} := \texttt{marg}(\Omega, v, x, \ell)$ be the value returned by our algorithm. We define

$$\mathcal{E}_{\Omega,v,\ell} := \max_{y \in [q]} \log\left(\frac{\hat{p}_{\Omega,v,y,\ell}}{\mathbf{Pr}_\Omega[c(v) = y]}\right)$$

with the convention $0/0 = 1$.

Let $\Omega = (G, \Lambda, \sigma)$ with $G = (V, E)$ be an instance of Potts model, $v \in V \setminus \Lambda$ be a vertex, $x \in [q]$ be a color and $\ell \in \mathbb{Z}$ be an integer. We can recursively identify each vertex in the computation tree of $\texttt{marg}(\Omega, v, x, \ell)$ with a subtree of $T = T_{\text{SAW}}(G, v)$:

- the root of the computation tree is identified with the root of T, i.e., the single vertex path (v);
- assuming the notations used in the description of Algorithm 2, if $\texttt{marg}(\Omega, v, x, \ell)$ is identified with a subtree of T rooted at self-avoiding walk P, then for every $\rho \in \mathcal{F}(B(v))$ and $i \in [m]$, the routine $\texttt{marg}(\Omega_i^\rho, v_i, \rho_i, \ell - P_i)$ is identified with the subtree of T rooted at the concatenation of P and P_i.

With this property, the following lemma relates our error of estimation to the error function defined before.

▶ **Lemma 17.** *Let $\Omega = (G, \Lambda, \sigma)$ with $G = (V, E)$ be an instance of Potts model, $v \in V \setminus \Lambda$ be a vertex, $x \in [q]$ be a color and $\ell \in \mathbb{Z}$ be an integer.*

Let S denote the set of vertices in T that can be identified to the leaves of the computation tree of `marg` (Ω, v, x, ℓ). *Let $T = T_{\mathrm{SAW}}(G, v)$, then we have*

$$\mathcal{E}_{\Omega,v,\ell} \le \mathcal{E}_{T,S}.$$

The key to prove Lemma 17 is to establish the *one-step contraction* of $\mathcal{E}_{\Omega,v,\ell}$, as stated in the following lemma:

▶ **Lemma 18.** *Let $\Omega = (G, \Lambda, \sigma)$ with $G = (V, E)$ be an instance of Potts model, $v \in V \setminus \Lambda$ be a vertex and $\ell \in \mathbb{Z}$ be an integer. Let $B = B(v)$ the minimal permissive block in G containing v. Assume the edge boundary $\delta B = \{u_i v_i \mid i \in [m]\}$ where $v_i \notin B$.*

Then it holds that

$$\mathcal{E}_{\Omega,v,\ell} \le \sum_{i=1}^{m} \frac{1-\beta}{q-1-(1-\beta)\deg_G(v_i)} \cdot \max_{\rho \in \mathcal{F}(B)} \mathcal{E}_{\Omega_i^\rho, v_i, \ell_i}$$

where Ω_i^ρ is defined in Section 3 and $\ell_i = \ell - |P_i|$ for the self-avoiding walk P_i chosen in Algorithm 2.

Proof. Let $\pi \in \mathcal{F}(B)$ be a coloring of the block B. We use $\hat{p}_{\Omega,B,\pi,\ell} =$ `marg-block` (Ω, B, π, ℓ) to denote the value return by our estimation algorithm for block marginals. Let x denote the color that achieves the maximum in the definition of $\mathcal{E}_{\Omega,v,\ell}$, we have

$$\begin{aligned}\mathcal{E}_{\Omega,v,\ell} &= \log\left(\frac{\hat{p}_{\Omega,v,x,\ell}}{\mathbf{Pr}_\Omega[c(v)=x]}\right) \le \log\left(\frac{\sum_{\substack{\pi\in\mathcal{F}(B)\\ s.t.\ \pi(v)=x}} \hat{p}_{\Omega,B,\pi,\ell}}{\sum_{\substack{\pi\in\mathcal{F}(B)\\ s.t.\ \pi(v)=x}} \mathbf{Pr}_\Omega[c(B)=\pi]}\right)\\ &\le \max_{\substack{\pi\in\mathcal{F}(B)\\ s.t.\ \pi(v)=x}} \log\left(\frac{\hat{p}_{\Omega,B,\pi,\ell}}{\mathbf{Pr}_\Omega[c(B)=\pi]}\right).\end{aligned}$$

By our algorithm, all the marginals in the recursion satisfies the upper bound in Lemma 11, it then follows from Lemma 10 that for every $\pi \in \mathcal{F}(B)$, it holds that

$$\log\left(\frac{\hat{p}_{\Omega,B,\pi,\ell}}{\mathbf{Pr}_\Omega[c(B)=\pi]}\right) \le \sum_{i=1}^{m} \frac{2(1-\beta)}{q-(1-\beta)\deg_G(v_i)-1} \cdot \max_{\rho\in\mathcal{F}(B)} \mathcal{E}_{\Omega_i^\rho, v_i, \ell_i}. \qquad ◀$$

We can use Lemma 18 to prove Lemma 17.

Proof of Lemma 17. We apply induction on $T := T_{\mathrm{SAW}}(G, v)$. The base case is that $v \in \Lambda$ or $v \in S$. If $v \in S$ and $\beta > 0$, then by Lemma 11 and Lemma 12, it holds that

$$\mathcal{E}_{\Omega,v,\ell} = \max_{y\in[q]} \log\left(\frac{\hat{p}_{\Omega,v,y,\ell}}{\mathbf{Pr}_\Omega[c(v)=y]}\right) \le q + n\log\frac{1}{\beta}.$$

If $v \in S$ and $\beta = 0$, by Proposition 8 and Lemma 11, we have

$$\mathcal{E}_{\Omega,v,\ell} = \max_{y\in[q]} \log\left(\frac{\hat{p}_{\Omega,v,y,\ell}}{\mathbf{Pr}_\Omega[c(v)=y]}\right) \le n\log q.$$

If $v \in \Lambda$, then $\mathcal{E}_{\Omega,v,\ell} = 0$.

Now assume $v \notin S \cup \Lambda$ and denote $B = B(v)$ the minimal permissive block containing v. Assume the edge boundary $\delta B = \{u_i v_i \mid u_i \in B\}$.

It then follows from Lemma 18 that

$$\mathcal{E}_{\Omega,v,\ell} \leq \sum_{i=1}^{m} \frac{1-\beta}{q-1-(1-\beta)\deg_G(v_i)} \cdot \max_{\rho\in\mathcal{F}(B)} \mathcal{E}_{\Omega_i^\rho,v_i,\ell_i}.$$

Recall for every $i \in [m]$, we define in Algorithm 2 a self-avoiding walk P_i containing $u_i v_i$ with every intermediate vertices in B. For every $u \in P_i$ such that $u \neq v, v_i$, it holds that $\delta(\deg_G(u)) = 1)$. With this property, if we use T_i to denote the subtree of T rooted at P_i, then

$$\sum_{i=1}^{m} \frac{1-\beta}{q-q-(1-\beta)\deg_G(v_i)} \cdot \mathcal{E}_{T_i,S} \leq \mathcal{E}_{T,S}.$$

We can then complete the proof by using induction hypothesis to show

$$\mathcal{E}_{\Omega_i^\rho,v_i,\ell_i} \leq \mathcal{E}_{T_i,S}$$

for every $i \in [m]$ and $\rho \in \mathcal{F}(B)$. ◀

We are now ready to prove the main lemma of this section.

Proof of Lemma 13

We can assume $v \notin \Lambda$, otherwise, the color on v is fixed by σ.

Let $T := T_{\mathrm{SAW}}(G, v)$. For every ℓ, let S_ℓ denote the set of vertices at which the procedure $\texttt{marg}(\Omega, v, x, \ell)$ terminates. Then it follows Lemma 17,

$$\log\left(\frac{\hat{p}_{\Omega,v,x,\ell}}{\mathbf{Pr}_\Omega[c(v)=x]}\right) \leq \mathcal{E}_{\Omega,v,\ell} \leq \mathcal{E}_{T,S_\ell}.$$

Note that $\mathrm{dist}_T(v, S_\ell) \geq \ell$, since $\delta(\cdot)$ is a contraction function for $\mathcal{G}$, we have $\mathcal{E}_{T,S_\ell} \leq n^C\gamma^\ell$ for some constants $C > 0$ and $0 < \gamma < 1$. This implies that for some constant $C_0 > 0$, if $\ell \geq C_0 \log n$, then

$$1 - O\left(\frac{1}{n^3}\right) \leq \frac{\hat{p}_{\Omega,v,x,\ell}}{\mathbf{Pr}_\Omega[c(v)=x]} \leq 1 + O\left(\frac{1}{n^3}\right).$$

To bound the running time of $\texttt{marg}(\Omega, v, x, \ell)$, we can apply Lemma 15 to conclude that if $\ell = \Theta(\log n)$, then the algorithm must terminate at depth $L = \Theta(\log n)$ of T, i.e., for every $u \in S_\ell$, it holds that $\mathrm{dist}_T(v, u) \leq L$.

We use T_ℓ to denote the subtree of T obtained by removing all descendants of S_ℓ and let $\mathcal{L}(T_\ell)$ to denote the set of self-avoiding walks corresponding to leaves of T_ℓ. Let $\tau_{\Omega,v,\ell}$ to denote the maximum running time of $\texttt{marg}(\Omega, v, x, \ell)$ over all colorings $x \in [q]$, we apply induction on T_ℓ to show that for some $C_1 > 0$, it holds that

$$\tau_{\Omega,v,\ell} \leq n^{C_1} \cdot \sum_{P\in\mathcal{L}(T_\ell)} q^{2|B(P)|}. \tag{4}$$

The base case is that $v \in S_\ell$ or $v \in \Lambda$ and our bound for running time trivially holds. Otherwise, denote $B = B_\Omega(v)$ the minimal permissive block containing v. Assume the edge boundary $\delta B = \{u_i v_i \mid u_i \in B\}$. We have for some constant $C_2 > 0$, it holds that

$$\tau_{\Omega,v,\ell} \leq q^{|B_\Omega(v)|} n^{C_2} + q^{|B_\Omega(v)|} \sum_{i=1}^{m} \max_{\rho\in\mathcal{F}(B(v))} \tau_{\Omega_i^\rho,v_i,\ell_i} \tag{5}$$

Recall that P_i is a self-avoiding walk from v to v_i containing u_i with every intermediate vertex in B. Let T_i denote the subtree of T rooted at P_i. Then we can apply induction hypothesis to obtain

$$\tau_{\Omega_i^\rho, v_i, \ell_i} \leq n^{C_1} \cdot \sum_{P \in \mathcal{L}(T_{\ell_i})} q^{2\left|B_{\Omega_i^\rho}(P)\right|} \tag{6}$$

for every $\rho \in \mathcal{F}(B)$ and $i \in [m]$. Furthermore, by our construction of Ω_i^ρ and the definition of permissive block, we have $|B_\Omega(v)| + \left|B_{\Omega_i^\rho}(v_i)\right| \leq \left|B_\Omega(v) \cup B_{\Omega_i^\rho(v_i)}(v_i)\right|$ for every $\rho \in \mathcal{F}(B)$ and $i \in [m]$. Plugging (6) into (5) proves (4).

Since G is locally sparse, we know that the term $q^{2|B(P)|}$ is bounded by a polynomial in n for every $P \in \mathcal{L}(T_\ell)$. It remains to show that $|\mathcal{L}(T_\ell)|$ is bounded by a polynomial in n. To see this, consider the contribution of a walk P in $\mathcal{L}(T_\ell)$ of length k to the quantity $\mathcal{E}_\delta(v,k)$ defined in (1). The contribution of this walk is at least $\frac{1}{\mathrm{poly}(n)}$ since $k = O(\log n)$ and for every $u \in P$, the value $\delta(u)$ is bounded below by a constant. It then follows from the fact that $\delta(\cdot)$ is a contraction function for G, there are at most polynomial many leaves in T_ℓ for our choice of ℓ.

5 The FPTAS and the sampling algorithm

In this section, we prove Theorem 7, by using the correlation decay property established in Section 4.

Proof of Theorem 7. Let $\Omega = (G, \varnothing, \varnothing)$ be an instance of Potts model, where $G = (V, E) \in \mathcal{G}$. Without loss of generality, we give an algorithm to compute an approximation of the partition function $\hat{Z}(\Omega)$ satisfying

$$1 - O\left(\frac{1}{n^2}\right) \leq \frac{\hat{Z}(\Omega)}{Z(\Omega)} \leq 1 + O\left(\frac{1}{n^2}\right).$$

Since our family of instances of Potts model is "self-embeddable" in the sense of [22], the algorithm can be boosted into an FPTAS.

Assume $V = \{v_1, \ldots, v_n\}$. First find a configuration $\sigma \in [q]^V$ such that $w_G(\sigma) > 0$. This task is trivial when $\beta > 0$. When $\beta = 0$, since G is q-colorable, we can also do it in polynomial time:

- If the graph is not empty, then choose a vertex v and find a feasible coloring of $B(v)$. Then remove $B(v)$ from the graph and repeat the process.

If G is q-colorable, then $G[V \setminus B(v)]$ is colorable as the boundary of $B(v)$ consists of low-degree vertices, thus the above process will end with a proper coloring of G, which is the union of colorings found at each step. The process terminates in polynomial time since $\mathcal{G}$ is locally sparse and thus the size of every $B(v)$ is $O(\log n)$.

With σ in hand, we have

$$\begin{aligned} Z(\Omega) = w_G(\sigma)/\mathbf{Pr}_\Omega\left[c(V) = \sigma\right] &= w_G(\sigma)\left(\mathbf{Pr}_\Omega\left[\bigwedge_{i=1}^n c(v_i) = \sigma(v_i)\right]\right)^{-1} \\ &= w_G(\sigma)\left(\prod_{i=1}^n \mathbf{Pr}_\Omega\left[c(v_i) = \sigma(v_i) \,\middle|\, \bigwedge_{j=1}^{i-1} c(v_j) = \sigma(v_j)\right]\right)^{-1} \end{aligned}$$

For every $i \in [n]$, let $\Omega_i = (G, \Lambda_i, \sigma_i)$ where $\Lambda_i = \{v_1, \dots, v_{i-1}\}$ and $\sigma_i(v_j) = \sigma(v_j)$ for every $j = 1, \dots, i-1$. We have

$$Z(\Omega) = w_G(\sigma)\left(\prod_{i=1}^{n} \mathbf{Pr}_{\Omega_i}\left[c(v_i) = \sigma(v_i)\right]\right)^{-1}.$$

Note that the graph class $\mathcal{G}$ is closed under the operation of fixing some vertex to a specific color, we can apply Lemma 13 for every Ω_i and obtain $\hat{p}_i$ such that

$$1 - O\left(\frac{1}{n^3}\right) \le \frac{\hat{p}_i}{\mathbf{Pr}_{\Omega_i}\left[c(v_i) = \sigma(v_i)\right]} \le 1 + O\left(\frac{1}{n^3}\right).$$

Let $\hat{Z}(\Omega) = w_G(\sigma)\left(\prod_{i=1}^{n} \hat{p}_i\right)^{-1}$, then Theorem 13 implies that

$$1 - O\left(\frac{1}{n^2}\right) \le \frac{\hat{Z}(\Omega)}{Z(\Omega)} \le 1 + O\left(\frac{1}{n^2}\right).$$

This approximate counting algorithm implies a sampling algorithm via Jerrum-Valiant-Vazirani reduction[17]. ◀

6 Random Graphs

In this section, we prove Theorem 2. We first prove the following properties of $\mathcal{G}(n, d/n)$.

▶ **Theorem 19.** *Let d be a sufficiently large constant, $q > 3(1-\beta)d + 4$ and $G = (V, E) \sim \mathcal{G}(n, d/n)$. Then with probability $1 - o(1)$, the following holds*

- *there exist two universal positive constants $C > 0, \gamma < 1$ such that $\mathcal{E}_\delta(v, \ell) < n^C \gamma^\ell$ for all $v \in V$ and for all $\ell = o(\sqrt{n})$, where $\mathcal{E}_\delta(v, \ell)$ is defined in (1);*
- *if $\beta = 0$, then G is q-colorable;*
- *there exists a universal constant $C > 0$ such that for every path P in G of length ℓ, $|B(P)| \le C(\ell + \log n)$.*

Note that the first property in above theorem impose an upper bound on ℓ. This is not harmful as our algorithms for FPTAS and sampling only require the property holds for $\ell = O(\log n)$. Thus Theorem 19 and Theorem 7 together imply Theorem 2.

It is well-known that when $\beta = 0$, G is q-colorable with high probability (see e.g., [12]), we verify the first property in Lemma 20 and the third property in Lemma 22.

6.1 Contraction function for random graphs

▶ **Lemma 20.** *Let $d > 1$, $0 \le \beta < 1$ and $q > 3(1-\beta)d + 4$ be constants. Let $G = (V, E) \sim \mathcal{G}(n, d/n)$. There exist two positive constants $C > 0$ and $\gamma < 1$ such that with probability $1 - O\left(\frac{1}{n}\right)$, for every $v \in V$ and every $\ell = o(\sqrt{n})$, it holds that*

$$\mathcal{E}_\delta(v, \ell) \le n^C \gamma^\ell$$

We first prove a technical lemma.

▶ **Lemma 21.** *Let $0 \le \beta < 1$ be a constant. Let $f_q(d) : \mathbb{R}^{\ge 0} \to \mathbb{R}^{\ge 0}$ be a piece wise function defined as*

$$f_q(d) := \begin{cases} \frac{2(1-\beta)}{q-1-(1-\beta)d} & \text{if } d \le \frac{q-1}{1-\beta} - 2 \\ 1 & \text{otherwise.} \end{cases}$$

Let X be a random variable distributed according to binomial distribution $\mathrm{Bin}(n, \frac{\Delta}{n})$ where $\Delta > 1$ is a constant. Then for $q \geq 3(1-\beta)\Delta + 2$ and all sufficiently large n, it holds that $\mathbf{E}\left[f_q(X)\right] < \frac{1}{\Delta}$.

Proof. Let $\lambda = 1 - \beta$. Since $f(d)$ is decreasing in q, we can assume $q = 3\lambda\Delta + 2$. Note that

$$\mathop{\mathbf{E}}_{d\sim\mathrm{Bin}\left(n,\frac{\Delta}{n}\right)}[f(d)] \leq \frac{1}{\Delta} \iff \mathop{\mathbf{E}}_{d\sim\mathrm{Bin}\left(n,\frac{\Delta}{n}\right)}[1-f(d)] \geq \frac{\Delta-1}{\Delta}.$$

Let $g(x) := 1 - f(x)$, then

$$\mathop{\mathbf{E}}_{d\sim\mathrm{Bin}\left(n,\frac{\Delta}{n}\right)}[1-f(d)] = \sum_{k=0}^{\lfloor\frac{q-1}{\lambda}-2\rfloor} g(k)\cdot p(k)$$

where $p(k) = \binom{n}{k}\left(\frac{\Delta}{n}\right)^k\left(1-\frac{\Delta}{n}\right)^{n-k}$.

Define

$$\begin{aligned}\tilde{g}(x) := 1 &- \frac{2\lambda}{q-1-\lambda\Delta} - \frac{2\lambda^2(x-\Delta)}{(q-1-\lambda\Delta)^2} - \frac{2\lambda^3(x-\Delta)^2}{(q-1-\lambda\Delta)^3} - \frac{2\lambda^4(x-\Delta)^3}{(q-1-\lambda\Delta)^4}\\ &- \frac{2\lambda^5(x-\Delta)^4}{(q-1-\lambda\Delta)^5} - \frac{2\lambda^6(x-\Delta)^5}{(q-1-\lambda\Delta)^6} - \frac{2\lambda^6(x-\Delta)^6}{(q-1-\lambda\Delta)^6}.\end{aligned}$$

Then

$$g(x) - \tilde{g}(x) = \frac{2\lambda^6(q-1-\lambda-x\lambda)(x-\Delta)^6}{(q-1-x\lambda)(q-1-\lambda\Delta)^6},$$

which is positive for $x \leq \lfloor\frac{q-1}{\lambda}-2\rfloor$.

We now prove that

$$\sum_{k=0}^{\lfloor\frac{q-1}{\lambda}-2\rfloor} \tilde{g}(k)\cdot p(k) \geq \frac{\Delta-1}{\Delta}.$$

The expectation of $\tilde{g}(k)$ can be computed directly:

$$\mathbf{E}\left[\tilde{g}(k)\right] = \frac{1}{n^5(q-1-\lambda\Delta)^6}\cdot\left(C_5n^5 + C_4n^4 \pm O(n^3)\right),$$

where

$$\begin{aligned}C_5 &= 1 - 2\lambda + (12\lambda - 20\lambda^2 - 2\lambda^3 - 2\lambda^4 - 2\lambda^5 - 4\lambda^6)\Delta\\ &\quad + (60\lambda^2 - 80\lambda^3 - 12\lambda^4 - 14\lambda^5 - 74\lambda^6)\Delta^2 + (160\lambda^3 - 160\lambda^4 - 24\lambda^5 - 50\lambda^6)\Delta^3\\ &\quad + (240\lambda^4 - 160\lambda^5 - 16\lambda^6)\Delta^4 + (192\lambda^5 - 64\lambda^6)\Delta^5 + 64\lambda^6\Delta^6;\\ C_4 &= 2\lambda^3(1 + 3\lambda + 7\lambda^2 + 46\lambda^3)\Delta^2 + 2\lambda^3(6\lambda + 18\lambda^2 + 234\lambda^3)\Delta^3\\ &\quad + 2\lambda^3(12\lambda^2 + 69\lambda^3)\Delta^4 + 16\lambda^6\Delta^5.\end{aligned}$$

Since $C_4 > 0$, thus for sufficiently large n, it holds that

$$\mathbf{E}\left[\tilde{g}(x)\right] \geq \frac{C_5}{(q-1-\lambda\Delta)^6}.$$

We also have that

$$\mathbf{E}\left[\tilde{g}(x)\right] = \sum_{k=0}^{\lfloor \frac{q-1}{\lambda}-2 \rfloor} \tilde{g}(k) \cdot p(k) + \sum_{k=\lfloor \frac{q-1}{\lambda}-1 \rfloor}^{n} \tilde{g}(k) \cdot p(k)$$

It can be verified that $\tilde{g}(x)$ is monotonically decreasing in x when $x \geq \frac{q-1}{\lambda} - 2$ and $\tilde{g}\left(\frac{q-1}{\lambda} - 2\right) = -\left(\frac{1+2\lambda(\Delta-1)}{1+2\lambda\Delta}\right)^6 < 0$.

Thus we have

$$\sum_{k=0}^{\lfloor \frac{q-1}{\lambda}-2 \rfloor} \tilde{g}(k) \cdot p(k) \geq \mathbf{E}\left[\tilde{g}(x)\right] \geq \frac{C_5}{(q-1-\lambda\Delta)^6} = \frac{\Delta - 1}{\Delta} + h(\Delta)$$

where

$$\begin{aligned} h(\Delta) = \big(&1 + 10\lambda\Delta + (40\lambda^2 - 2\lambda^3 - 2\lambda^4 - 2\lambda^5 - 4\lambda^6)\Delta^2 \\ &+(80\lambda^3 - 12\lambda^4 - 14\lambda^5 - 74\lambda^6)\Delta^3 + (80\lambda^4 - 24\lambda^5 - 50\lambda^6)\Delta^4\big) \\ &+(32\lambda^5 - 16\lambda^6)\Delta^5\big) \cdot \left(\Delta(1+2\lambda\Delta)^6\right)^{-1}. \end{aligned}$$

It can be verified that $h(\Delta)$ is positive for every $0 < \lambda < 1$ and $\Delta \geq 1$. ◀

Proof of Lemma 20. Let $v \in V$ be arbitrary fixed and $T_v = T_{\mathrm{SAW}}(G, v)$ and $\ell > 0$ be an integer. By linearity of expectation, we have

$$\mathbf{E}\left[\mathcal{E}_\delta(v,\ell)\right] \leq n^\ell \left(\frac{d}{n}\right)^\ell \mathbf{E}\left[\prod_{i=1}^{\ell} \delta(\deg_G(v_i)) \,\middle|\, P = (v, v_1, \ldots, v_\ell) \text{ is a path}\right].$$

Fix a tuple $P = (v, v_1, \ldots, v_\ell)$. To calculate the expectation, we construct an independent sequence whose product dominates $\prod_{i=1}^{\ell} \delta(\deg_G(v_i))$ as follows.

Conditioning on $P = (v, v_1, \ldots, v_\ell)$ being a path in G. Let $X_1, X_2, \ldots, X_\ell$ be random variables such that each X_i represents the number of edges between v_i and vertices in $V \setminus \{v_1, \ldots, v_\ell\}$; and let Y be a random variable representing the number of edges between vertices in $\{v_1, \ldots, v_\ell\}$ except for the edges in the path $P = (v, v_1, \ldots, v_\ell)$. Then $X_1, \ldots, X_\ell, Y$ are mutually independent binomial random variables with each X_i distributed according to $\mathrm{Bin}(n-\ell, \frac{d}{n})$ and Y distributed according to $\mathrm{Bin}(\binom{\ell}{2} - \ell + 1, \frac{d}{n})$, and for each v_i in the path we have $\deg_G(v_i) = X_i + 2 + Y_i$ with some $Y_1 + Y_2 + \cdots + Y_\ell = 2Y$.

Note that $\delta(\deg_G(v_i)) = f_q(\deg_G(v_i))$ where the function $f_q(x)$ is defined in Lemma 21. Note that the ratio $f_q(x)/f_q(x-1)$ is always upper bounded by 2, and we have $f_q(x+1) \leq f_{q-1}(x)$. Thus, conditioning on that $P = (v, v_1, \ldots, v_\ell)$ is a path, the product $\prod_{i=1}^{\ell} \delta_{q,\beta}(\deg_G(v_i))$ can be bounded as follows:

$$\prod_{i=1}^{\ell} \delta(\deg_G(v_i)) = \prod_{i=1}^{\ell} f_q(X_i + Y_i + 2) \leq 2^{2Y} \prod_{i=1}^{\ell} f_{q-2}(X_i).$$

Let $d' = \frac{q-4}{3(1-\beta)}$, then we have $d' > d$. Let X be a binomial random variable distributed according to $\mathrm{Bin}(n, \frac{d'}{n})$, thus X probabilistically dominates every X_i whose distribution is $\mathrm{Bin}(n-\ell, \frac{d}{n})$. Since $X_1, X_2, \ldots, X_\ell, Y$ are mutually independent conditioning on $P = (v, v_1, \ldots, v_\ell)$ being a path in G, for any $P = (v, v_1, \ldots, v_\ell)$ we have

$$\mathbf{E}\left[\prod_{i=1}^{\ell} \delta(\deg_G(v_i)) \,\middle|\, P \text{ is a path}\right] \leq \mathbf{E}\left[4^Y \prod_{i=1}^{\ell} f_{q-2}(X_i)\right] \leq \mathbf{E}\left[4^Y\right] \mathbf{E}\left[f_{q-2}(X)\right]^\ell.$$

Recall that $Y \sim \mathrm{Bin}\left(\binom{\ell}{2} - \ell + 1, \frac{d}{n}\right)$, the expectation $\mathbf{E}\left[4^Y\right]$ can be bounded as

$$\mathbf{E}\left[4^Y\right] \le \sum_{k=0}^{\ell^2} 4^k \binom{\ell^2}{k} \left(\frac{d}{n}\right)^k \left(1 - \frac{d}{n}\right)^{\ell^2 - k} = \left(1 + \frac{3d}{n}\right)^{\ell^2} \le \exp\left(\frac{3d\ell^2}{n}\right).$$

Since $q - 2 \ge 3(1-\beta)d' + 2$, it follows from Lemma 21 that $\mathbf{E}\left[f_{q-2}(X)\right] \le \frac{1}{d'} = \frac{3(1-\beta)}{q-4}$. Therefore,

$$\begin{aligned}\mathbf{E}\left[\prod_{i=1}^{\ell} \delta(\deg_G(v_i)) \;\middle|\; P \text{ is a path}\right] &\le \exp\left(\frac{3d\ell^2}{n}\right)\left(\frac{3(1-\beta)}{q-4}\right)^{\ell} \\ &\le \frac{1}{d^\ell} \cdot \exp\left(-\ell \log\left(\frac{q-4}{3d(1-\beta)}\right) + \frac{3d\ell^2}{n}\right).\end{aligned}$$

Since $\ell = o(\sqrt{n})$,

$$\mathbf{E}\left[\mathcal{E}_\delta(v, \ell)\right] \le \exp\left(-\ell \log\left(\frac{q-4}{3d(1-\beta)}\right) + o(1)\right).$$

Then the lemma follows from the Markov inequality and the union bound. ◀

6.2 Locally sparsity for random graph

▶ **Lemma 22.** *Let $\varepsilon > 0$ be some fixed constant. Let d be a sufficiently large number, $q \ge (2+\varepsilon)d$ and $0 \le \beta < 1$ be constants. Let $G = (V, E) \sim \mathcal{G}(n, d/n)$. There exists a constant $C > 0$ such that with probability $1 - O\left(\frac{1}{n}\right)$, for every path P in G of length ℓ, $|B(P)| \le C(\ell + \log n)$.*

Given $P = (v_1, \ldots, v_L)$, we are going to upper bound the probability

$$\mathbf{Pr}\left[|B(P)| \ge t \mid P \text{ is a path}\right] \tag{7}$$

for every $t > 0$.

A vertex v is a high-degree vertex if $\deg_G(v) \ge \frac{q-1}{1-\beta} - 2$. Thus the probability (7) is maximized when $\beta = 0$. Note that conditioning on P is a path gives each vertex at most two degrees, we can redefine the notion of "high-degree" as $\deg_G(v) \ge q - 5$ and drop the condition that P is a path. Thus it is sufficient to upper bound

$$\mathbf{Pr}\left[|B(P)| \ge t\right]$$

with our new definition of high-degree vertices.

Let $G = (V, E)$ be a graph. We now describe a BFS procedure to generate $B^*(P) := B(P) \cup \partial B(P)$. Since $B^*(P)$ is always a superset of $B(P)$, it is sufficient to bound $\mathbf{Pr}\left[|B^*(P)| \ge t\right]$. For a vertex $v \in V$, we use $N_G(v)$ to denote the set of neighbors of v in G.

Initially, we have a counter $i = 0$, a graph $G_0 = G$, a set of active vertices $\mathcal{A}_0 = \{v_1, v_2, \ldots, v_L\}$ and a set of used vertices $\mathcal{U}_0 = \varnothing$.

(P1)

1. Increase the counter i by one.
2. (If $i \le L$) Define $G_i(V_i, E_i) = G_{i-1}[V_{i-1} \setminus \{v_i\}]$. Let $\mathcal{U}_i = \mathcal{U}_{i-1} \cup \{v_i\}$. Let $\mathcal{A}_i = (\mathcal{A}_{i-1} \cup N_{G_{i-1}}(v_i)) \setminus \mathcal{U}_i$. Goto 1.

3. (If $i > L$) Terminate if $\mathcal{A}_{i-1} = \varnothing$. Otherwise, let $u \in \mathcal{A}_{i-1}$ and let $\mathcal{U}_i = \mathcal{U}_{i-1} \cup \{u\}$.
   a. (If $|N_G(u)| \geq q-5$) Define $G_i(V_i, E_i) = G_{i-1}[V_{i-1} \setminus \{u\}]$. Let $\mathcal{A}_i = (\mathcal{A}_{i-1} \cup N_{G_{i-1}}(v_i)) \setminus \mathcal{U}_i$. Goto 1.
   b. (If $|N_G(u)| < q-5$) Define $G_i = G_{i-1}$. Let $\mathcal{A}_i = \mathcal{A}_{i-1} \setminus \mathcal{U}_i$. Goto 1

The following proposition is immediate:

▶ **Proposition 23.** *Assume the algorithm terminates at step t, then $B^*(P) = \mathcal{U}_{t-1}$ and $|B^*(P)| = t-1$.*

Let $R = \{r_1, r_2, \ldots, r_L\}$ be a set and each r_i is the root of tree T_i. We now describe a BFS procedure to explore these L trees. For a vertex v, we use $C(v)$ to denote its children.

Initially, we have a counter $i = 0$ and a set of active vertices $\mathcal{B}_0 = R$.

(P2)

1. Increase the counter i by one.
2. (If $i \leq L$) Let $\mathcal{B}_i = (\mathcal{B}_{i-1} \cup C(r_i)) \setminus \{r_i\}$. Goto 1.
3. (If $i > L$) Terminate if $\mathcal{B}_{i-1} = \varnothing$. Otherwise, let $w \in \mathcal{B}_{i-1}$
   a. (If $|C(w)| \geq \frac{q-5}{2}$) Let $\mathcal{B}_i = (\mathcal{B}_{i-1} \cup C(u)) \setminus \{w\}$. Goto 1.
   b. (If $|C(w)| < \frac{q-5}{2}$) Let $\mathcal{B}_i = \mathcal{B}_{i-1} \setminus \{w\}$. Goto 1.

Now assume $G \sim \mathcal{G}(n, d/n)$ and for every $i \in [L]$, T_i is a branching process with distribution $\mathrm{Bin}(n, d/n)$, i.e., each $C(u) \sim \mathrm{Bin}([n], d/n)$. We can implement the (P1) when at each step i, the vertex u chosen from the active set sample its neighbors $N_{G_{i-1}}(u)$ according to $\mathrm{Bin}(V_i, d/n)$. This random process can be coupled with $\mathcal{G}(n, d/n)$ such that $B^*(P)$ found by it is always a superset of the one in $\mathcal{G}(n, d/n)$.

We now construct a coupling of (P1) and (P2) with the property that the later one always terminates no earlier than the former one.

At each step $i \geq 1$, let u and w be the vertex chosen from $\mathcal{A}_i$ and $\mathcal{B}_i$ respectively ($u = v_i$ and $w = r_i$ if $i \leq L$). Then $\left|N_{G_{i-1}}(u)\right| \sim \mathrm{Bin}(|V_i|, d/n)$. We couple it with some $x \sim \mathrm{Bin}(n, d/n)$ with the property that $x \geq \left|N_{G_{i-1}}(v_i)\right|$ and let $C(w)$ be a set with x elements.

▶ **Lemma 24.** *For every $i \geq 0$, the following two properties hold:*

(i1) *There exists a surjective mapping F_i from $\mathcal{B}_i$ to $\mathcal{A}_i$ in each step i.*

(i2) *For every $u \in \mathcal{A}_i$, we use $n_i(u)$ to denote the number of $w \in \mathcal{B}_i$ such that $F_i(w) = u$. Then for every $u \in \mathcal{A}_i$, $n_i(u) \geq |N_G(u)| - |N_{G_i}(u)|$.*

Proof. We apply induction on i to prove the lemma.

When $i = 0$, we let $F_0 : \mathcal{B}_0 \to \mathcal{A}_0$ be the function that $F_0(r_j) = v_j$ for every $j \in [L]$. Then both properties hold trivially.

Assume the lemma holds for smaller i. If $i \leq L$, since by our coupling, $|C(r_i)| \geq \left|N_{G_{i-1}}(v_i)\right|$, we can construct F_i by extending F_{i-1} with an arbitrary surjective mapping from $C(r_i)$ to $N_{G_{i-1}}(v_i)$. For every $u' \in \mathcal{A}_i$, if $u' \in N_{G_{i-1}}(v_i)$, then $n_i(u') \geq n_{i-1}(u') + 1$ and $\left|N_{G_{i-1}(u)}\right| - |N_{G_i}(u')| = 1$; otherwise $n_i(u') = n_{i-1}(u')$ and $\left|N_{G_{i-1}}(u)\right| = |N_{G_i}(u')|$. Induction hypothesis implies both $(i1)$ and $(i2)$ hold.

If $i > L$, we have to distinguish between cases:

- (If $|N_G(u)| \geq q-5$ and $N_{G_{i-1}}(u) \geq \frac{q-5}{2}$) We construct F_i by extending F_{i-1} with an arbitrary surjective mapping from $C(w)$ to $N_{G_{i-1}}(u)$, the same argument as $i \leq L$ case proves $(i1)$ and $(i2)$.

- (If $|N_G(u)| \geq q-5$ and $N_{G_{i-1}}(u) < \frac{q-5}{2}$) In this case, by induction hypothesis, we know that

$$n_{i-1}(u) \geq |N_G(u)| - \left|N_{G_{i-1}}(u)\right| \geq \frac{q-5}{2} > N_{G_{i-1}}(u).$$

Choose a surjective f from $F_{i-1}^{-1}(u)$ to $N_{G_{i-1}}(u)$ and construct F_i from F_{i-1} by replacing the mapping on $F_{i-1}^{-1}(u)$ by f. This is safe since $u \notin \mathcal{A}_i$. The same argument as before proves $(i1)$ and $(i2)$.
- (If $|N_G(u)| < q-5$) Construct $F_i = F_{i-1}$. Since everything does not change, the induction hypothesis implies $(i1)$ and $(i2)$. ◀

The first property above guarantees that (P2) terminates no earlier than (P1) and thus its stopping time is an upper bound for the size of $B^*(P)$ found by (P1).

(P2) can be modeled as follows:

1. Let $X \sim \mathrm{Bin}(n, d/n)$ and $X_1, X_2 \ldots$ be an infinite sequence of independent random variables defined as follows
   - For $i = 1, 2, \ldots, L$, X_i is an independent copy of X;
   - For $i > L$, X_i has following distribution

$$X_i = \begin{cases} 0 & \text{if } X < (q-5)/2 \\ X & \text{otherwise.} \end{cases}$$

2. $Y_1, Y_2, \ldots$ is an infinite sequence of random variables that $Y_0 = L$ and $Y_i = Y_{i-1} + X_i - 1$ for every $i \geq 1$.
3. $Z = \min_t \{Y_t = 0\}$.

The above process is identical to (P2), thus we have

▶ **Proposition 25.** (P2) *terminates after step t if and only if $Z > t$.*

Note that $Z > t$ implies $Y_t \geq 0$, we turn to bound the latter.

▶ **Lemma 26.** *There exist two constants $C_1, C_2 > 0$ depending on d and ε such that*

$$\mathbf{Pr}\left[Y_t \geq 0\right] \leq \exp\left(-C_1 t + C_2 L\right).$$

Proof. By the definition, $Y_{t+L} = L - (t+L) + \sum_{i=1}^{L+t} X_i = -t + \sum_{i=1}^{L} X_i + \sum_{i=L+1}^{L+t} X_i$. We know the distribution of X_is and we now compute their moment generating function. For every $s > 0$, it holds that

$$\mathbf{Pr}\left[Y_{t+L} \geq 0\right] = \mathbf{Pr}\left[e^{sY_{t+L}} \geq 1\right] \leq \mathbf{E}\left[e^{sY_{t+L}}\right] = e^{-st}\left(\mathbf{E}\left[e^{sX}\right]\right)^L \left(\mathbf{E}\left[e^{sX_{L+1}}\right]\right)^t.$$

Recall that $X \sim \mathrm{Bin}(n, d/n)$, we have $\mathbf{E}\left[e^{sX}\right] = \left(1 + \frac{d}{n}(e^s - 1)\right)^n \leq e^{d(e^s-1)}$. Let $p = (q-5)/2$, we have

$$\begin{aligned}
\mathbf{E}\left[e^{sX_{L+1}}\right] &= \mathbf{Pr}\left[X < p\right] + \sum_{k=\lfloor p \rfloor}^{n} e^{sk} \cdot \mathbf{Pr}\left[X = k\right] \\
&\leq 1 + \sum_{k=\lfloor p \rfloor}^{n} e^{sk} \cdot \mathbf{Pr}\left[X \geq k\right] \\
&\leq \exp\left(\sum_{k=\lfloor p \rfloor}^{\infty} e^{sk} \cdot \mathbf{Pr}\left[X \geq k\right]\right)
\end{aligned}$$

By Chernoff bound, for sufficiently large d, we have for some choices of $s > 0$ and $C_1 > 0$,

$$\sum_{k=\lfloor p \rfloor}^{\infty} e^{sk} \cdot \mathbf{Pr}\left[X \geq k\right] - s < -C_1'.$$

Let $C_2' = d(e^s - 1)$, we have

$$\mathbf{Pr}\left[Y_{t+L} \geq 0\right] \leq \exp\left(-C_1't + C_2'L\right).$$

This implies for some constants $C_1, C_2 > 0$,

$$\mathbf{Pr}\left[Y_t \geq 0\right] \leq \exp\left(-C_1t + C_2L\right). \blacktriangleleft$$

Proof of Lemma 22. By Lemma 26 and the union bound, the probability that there exists a path P in G of length ℓ such that $|B(P)| \geq t$ is upper bounded by

$$n \cdot n^\ell \cdot \left(\frac{d}{n}\right)^\ell \cdot \mathbf{Pr}\left[Y_t \geq 0\right] \leq n \cdot d^\ell \cdot \exp\left(-C_1t + C_2\ell\right) = O\left(\frac{1}{n}\right)$$

for $t = C(\ell + \log n)$ and sufficiently large constant C. ◀

Acknowledgement. This work was done in part while both authors were visiting the Simons Institute for the Theory of Computing.

References

1 Russ Bubley and Martin Dyer. Path coupling: A technique for proving rapid mixing in Markov chains. In *Proceedings of the 38th Annual IEEE Symposium on Foundations of Computer Science (FOCS'97)*, pages 223–231. IEEE, 1997.

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

3 Martin Dyer and Alan Frieze. Randomly coloring graphs with lower bounds on girth and maximum degree. *Random Structures & Algorithms*, 23(2):167–179, 2003.

4 Martin Dyer, Alan Frieze, Thomas Hayes, and Eric Vigoda. Randomly coloring constant degree graphs. *Random Structures & Algorithms*, 43(2):181–200, 2013.

5 Charilaos Efthymiou. A simple algorithm for random colouring $G(n, d/n)$ using $(2+\varepsilon)$d colours. In *Proceedings of the 23th Annual ACM-SIAM symposium on Discrete Algorithms (SODA'12)*, pages 272–280. SIAM, 2012.

6 Charilaos Efthymiou. Mcmc sampling colourings and independent sets of $G(n, d/n)$ near uniqueness threshold. In *Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'14)*, pages 305–316. SIAM, 2014.

7 Charilaos Efthymiou. Switching colouring of $G(n, d/n)$ for sampling up to Gibbs uniqueness threshold. In *In Proceedings of the 22nd European Symposium on Algorithms (ESA'14)*, pages 371–381. Springer, 2014.

8 Charilaos Efthymiou and Paul Spirakis. Random sampling of colourings of sparse random graphs with a constant number of colours. *Theoretical Computer Science*, 407(1):134–154, 2008.

9 Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Inapproximability for antiferromagnetic spin systems in the tree nonuniqueness region. *Journal of the ACM (JACM)*, 62(6):50, 2015.

10 David Gamarnik and Dmitriy Katz. Correlation decay and deterministic FPTAS for counting colorings of a graph. *Journal of Discrete Algorithms*, 12:29–47, 2012.

11 David Gamarnik, Dmitriy Katz, and Sidhant Misra. Strong spatial mixing of list coloring of graphs. *Random Structures & Algorithms*, 46(4):599–613, 2015.

12 Geoffrey Grimmett and Colin McDiarmid. On colouring random graphs. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 77, pages 313–324. Cambridge Univ Press, 1975.

13 Thomas Hayes. Randomly coloring graphs of girth at least five. In *Proceedings of the 35th Annual ACM Symposium on Symposium on Theory of Computing (STOC'03)*, pages 269–278. ACM, 2003.

14 Thomas Hayes and Eric Vigoda. A non-markovian coupling for randomly sampling colorings. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science (FOCS'03)*, pages 618–627. IEEE, 2003.

15 Thomas Hayes and Eric Vigoda. Coupling with the stationary distribution and improved sampling for colorings and independent sets. *The Annals of Applied Probability*, 16(3):1297–1318, 2006.

16 Mark Jerrum. A very simple algorithm for estimating the number of k-colorings of a low-degree graph. *Random Structures and Algorithms*, 7(2):157–166, 1995.

17 Mark Jerrum, Leslie Valiant, and Vijay Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoretical Computer Science*, 43:169–188, 1986.

18 Johan Jonasson. Uniqueness of uniform random colorings of regular trees. *Statistics & Probability Letters*, 57(3):243–248, 2002.

19 Pinyan Lu and Yitong Yin. Improved FPTAS for multi-spin systems. In *Proceedings of APPROX-RANDOM*, pages 639–654. Springer, 2013.

20 Michael Molloy. The Glauber dynamics on colorings of a graph with high girth and maximum degree. *SIAM Journal on Computing*, 33(3):721–737, 2004.

21 Elchanan Mossel and Allan Sly. Gibbs rapidly samples colorings of $G(n, d/n)$. *Probability theory and related fields*, 148(1-2):37–69, 2010.

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

23 Alistair Sinclair, Piyush Srivastava, Daniel Štefankovič, and Yitong Yin. Spatial mixing and the connective constant: Optimal bounds. In *Proceedings of the 26th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'15)*, pages 1549–1563. SIAM, 2015.

24 Alistair Sinclair, Piyush Srivastava, and Yitong Yin. Spatial mixing and approximation algorithms for graphs with bounded connective constant. In *Proceedings of the 54th Annual IEEE Symposium on Foundations of Computer Science (FOCS'13)*, pages 300–309. IEEE, 2013.

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

26 Yitong Yin. Spatial mixing of coloring random graphs. In *Proceedings of the 41st International Colloquium on Automata, Languages and Programming (ICALP'14, Track A)*, pages 1075–1086. Springer, 2014.