



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

What can be sampled locally?

Citation for published version:

Feng, W, Sun, Y & Yin, Y 2020, 'What can be sampled locally?', *Distributed Computing*, vol. 33, no. 3, pp. 227-253. <https://doi.org/10.1007/s00446-018-0332-8>

Digital Object Identifier (DOI):

[10.1007/s00446-018-0332-8](https://doi.org/10.1007/s00446-018-0332-8)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Distributed Computing

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.





What can be sampled locally?

Weiming Feng¹ · Yuxin Sun² · Yitong Yin³

Received: 8 November 2017 / Accepted: 17 April 2018 / Published online: 27 April 2018
© The Author(s) 2018

Abstract

The local computation of Linial [FOCS'87] and Naor and Stockmeyer [STOC'93] studies whether a locally defined distributed computing problem is locally solvable. In classic local computation tasks, the goal of distributed algorithms is to construct a feasible solution for some constraint satisfaction problem (CSP) locally defined on the network. In this paper, we consider the problem of sampling a uniform CSP solution by distributed algorithms in the LOCAL model, and ask whether a locally definable joint distribution is locally sample-able. We use Markov random fields and Gibbs distributions to model locally definable joint distributions. We give two distributed algorithms based on Markov chains, called *LubyGlauber* and *LocalMetropolis*, which we believe to represent two basic approaches for distributed Gibbs sampling. The algorithms achieve respective mixing times $O(\Delta \log n)$ and $O(\log n)$ under typical mixing conditions, where n is the number of vertices and Δ is the maximum degree of the graph. We show that the time bound $\Theta(\log n)$ is optimal for distributed sampling. We also show a strong $\Omega(\text{diam})$ lower bound: in particular for sampling independent set in graphs with maximum degree $\Delta \geq 6$. This gives a strong separation between sampling and constructing locally checkable labelings.

Keywords Distributed sampling algorithms · Local computation · LOCAL model · Gibbs sampling · Markov chain Monte Carlo

1 Introduction

Local computation and the LOCAL model. Locality of computation is a central theme in the theory of distributed computing. In the seminal works of Linial [44], and Naor and Stockmeyer [49], the locality of distributed computation and the locally definable distributed computing problems are respectively captured by the LOCAL model and the notion

of *locally checkable labeling (LCL)* problems. In the LOCAL model [49,52], a network of n processors is represented as an undirected graph, where each vertex represents a processor and each edge represents a bidirectional communication channel. Computations and communications are organized in synchronized rounds. In each round, each processor may receive a message of arbitrary size from each of its neighbors, perform an arbitrary local computation with the information collected so far, and send a message of arbitrary size to each of its neighbors. The output value for each vertex in a t -round protocol is determined by the local information within the t -neighborhood of the vertex. The local computation tasks are usually formulated as *labeling* problems, such as the locally checkable labeling (LCL) problems introduced in [49], in which the distributed algorithm is asked to construct a feasible solution of a constraint satisfaction problem (CSP) defined by local constraints with constant diameter in the network. Many problems can be expressed in this way, including various vertex/edge colorings, or local optimizations such as maximal independent set (MIS) and maximal matching.

A classic question for local computation is whether a locally definable problem is locally computable. Mathematically, this asks whether a feasible solution for a given local

A preliminary version of this article appeared in Proceedings of the 36th ACM Symposium on Principles of Distributed Computing (PODC).

✉ Yitong Yin
yinyt@nju.edu.cn
Weiming Feng
fengwm@smail.nju.edu.cn
Yuxin Sun
yxsun@cs.wisc.edu

- ¹ Department of Computer Science and Technology, Nanjing University, Nanjing, China
- ² Department of Computer Sciences, University of Wisconsin-Madison, Madison, WI, US
- ³ State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China

CSP can be *constructed* using only local information. There is a substantial body of research works dedicated to this question [2–5,10,24,25,29–31,34,39–42,44,49,54].

The local sampling problem Given an LCL problem which defines a local CSP on the network, aside from constructing a feasible solution of the local CSP, another interesting problem is to *sample* a *uniform* random feasible solution, e.g. to sample a uniform random proper coloring of the network G with a given number of colors. More abstractly, given an instance of local CSP which, say, treats the vertices in the network $G(V, E)$ as variables, a joint distribution of uniform random feasible solution $\mathbf{X} = (X_v)_{v \in V}$ is accordingly defined by these local constraints. Our main question is *whether a locally definable joint distribution can be sampled from locally*.

Intuitively, sampling could be substantially more difficult than labeling, because to sample a feasible solution is at least as difficult as to construct one, and furthermore, the marginal distribution of each random variable X_v in a jointly distributed feasible solution $\mathbf{X} = (X_v)_{v \in V}$ may already encapsulate certain amount of non-local information about the solution space.

Retrieving such information about the solution space (as in sampling) instead of constructing one solution (as in labeling) by distributed algorithms is especially well motivated in the context of distributed machine learning [14,15,17,32,50,57,61–63], where the data (the description of the joint distribution) is usually distributed among a large number of servers.

Besides uniform distributions, it is also natural to consider sampling from general non-uniform distributions over the solution space, which are usually formulated as graphical models known as the weighted CSPs [7], also known as factor graphs [47]. In this model, a probability distribution called the *Gibbs distribution* is defined over the space $\Omega = [q]^V$ of configurations, in such a way that each constraint of the weighted CSP contributes a nonnegative factor in the probability measure of a configuration in Ω . Due to Hammersley-Clifford's fundamental theorem [47, Theorem 9.3] of random fields, this model is universal for conditional independent (spatial Markovian) [47, Proposition 9.2] joint distributions. The conditional independence property roughly says that fixed a separator $S \subset V$ whose removal "disconnects" the variable sets A and B , given any feasible configuration $X_S = \sigma_S$ over S , the configurations X_A over A and X_B over B are conditionally independent.

We are particularly interested in a basic class of weighted local CSPs, namely the *Markov random fields (MRFs)*, where every local constraint (factor) is either a binary constraint over an edge or a unary constraint on a vertex. Specifically, given a graph $G(V, E)$ and a finite domain $[q] = \{1, 2, \dots, q\}$, the probability measure $\mu(\sigma)$ of each configuration $\sigma \in [q]^V$ under the Gibbs distribution μ is defined

to be proportional to the weight:

$$w(\sigma) := \prod_{e=uv \in E} A_e(\sigma_u, \sigma_v) \prod_{v \in V} b_v(\sigma_v), \quad (1)$$

where $\{A_e \in \mathbb{R}_{\geq 0}^{q \times q}\}_{e \in E}$ are non-negative $q \times q$ symmetric matrices and $\{b_v \in \mathbb{R}_{\geq 0}^q\}_{v \in V}$ are non-negative q -vectors, both specified by the instance of MRF. Examples of MRFs include combinatorial models such as independent set, vertex cover, graph coloring, and graph homomorphism, or physical models such as hardcore gas model, Ising model, Potts model, and general spin systems.

1.1 Our results

We give two Markov chain based distributed algorithms for sampling from Gibbs distributions. Given any $\epsilon > 0$, each algorithm returns a random output which is within total variation distance ϵ from the Gibbs distribution. Our expositions mainly focus on MRFs, although both algorithms can be extended straightforwardly to general weighted local CSPs.

In classic single-site Markov chains for Gibbs sampling, such as the Glauber dynamics, at each step a variable is picked at random and is updated according to its neighbors' current states. A generic approach for parallelizing a single-site sequential Markov chain is to update a set of non-adjacent vertices in parallel at each step. This natural idea has been considered in [32], also in a much broader context such as parallel job scheduling [12] or distributed Lovász local lemma [11,48]. For sampling from locally defined joint distributions, it is especially suitable because of the *conditional independence* property of MRFs.

Our first algorithm, named *LubyGlauber*, naturally parallelizes the Glauber dynamics by parallel updating vertices from independent sets generated by the "Luby step" in Luby's algorithm [1,46]. It is well known that Glauber dynamics achieves the mixing rate $\tau(\epsilon) = O(n \log(\frac{n}{\epsilon}))$ under the *Dobrushin's condition* for the decay of correlation [16,35]. By a standard coupling argument, the *LubyGlauber* algorithm achieves a mixing rate $\tau(\epsilon) = O(\Delta \log(\frac{n}{\epsilon}))$ under the same condition, where Δ is the maximum degree of the network. In particular, for uniform proper q -colorings, this implies:

Theorem 1 *If $q \geq \alpha \Delta$ for an arbitrary constant $\alpha > 2$, there is an algorithm which samples a uniform proper q -coloring within total variation distance $\epsilon > 0$ within $O(\Delta \log(\frac{n}{\epsilon}))$ rounds of communications on any graph $G(V, E)$ with $n = |V|$ vertices and maximum degree Δ , where Δ may be unbounded.*

A barrier for this natural approach is that it will perform poorly on general graphs with large chromatic number. The situation motivates us to ask following questions:

- Is it possible to update all variables in $X = (X_v)_{v \in V}$ simultaneously and still converge to the correct stationary distribution μ ?
- More concretely, is it always possible to sample almost uniform proper q -coloring, for a $q = O(\Delta)$, on any graphs $G(V, E)$ with $n = |V|$ vertices and maximum degree Δ , within $O(\log n)$ rounds of communications, especially when Δ is unbounded?

Surprisingly, the answers to both questions are “yes”. We give an algorithm, called the *LocalMetropolis* algorithm, achieving these goals. This is a bit surprising, since it seems to fully parallelize a process which is intrinsically sequential due to the massive local dependencies, especially on graphs with unbounded maximum degree. The algorithm follows the *Metropolis-Hastings* paradigm: at each step, it proposes to update all variables independently and then applies proper local filtrations to the proposals to ensure its convergence to the correct joint distribution. Our main discovery is that for locally defined joint distributions, the Metropolis filters are *localizable*.

The *LocalMetropolis* algorithm always converges to the correct Gibbs distribution. The analysis of its mixing time is more involved. In particular, for uniformly sampling proper q -coloring we show:

Theorem 2 *If $q \geq \alpha \Delta$ for an arbitrary constant $\alpha > 2 + \sqrt{2}$, there is an algorithm for sampling uniform proper q -coloring within total variation distance $\epsilon > 0$ in $O(\log(\frac{n}{\epsilon}))$ rounds of communications on any graph $G(V, E)$ with $n = |V|$ vertices and maximum degree at most $\Delta \geq 9$, where Δ may be unbounded.*

Neither of the algorithms abuses the power of the LOCAL model: each message is of $O(\log n)$ bits if the domain size $q = \text{poly}(n)$.

Due to the exponential correlation between variables in Gibbs distributions, the $O(\log(\frac{n}{\epsilon}))$ time bound achieved in Theorem 2 is optimal.

After the submission of this paper, two independent works [21,23] give the same distributed algorithm for sampling random q -coloring, which improves the *LocalMetropolis* algorithm by introducing a step of laziness as distributed symmetry breaking. This new algorithm achieves an $O(\log n)$ mixing time under the Dobrushin’s condition $q \geq (2 + \delta)\Delta$. Furthermore, for graphs with sufficiently large maximum degree and girth at least 9, it achieves an $O(\log n)$ mixing time when $q \geq (\alpha^* + \delta)\Delta$, where $\alpha^* \approx 1.763$ is the positive root of equation $x = e^{1/x}$. Another non-MCMC algorithm named distributed JVV sampler is given in [22], which successfully samples. For many locally definable joint distributions, this algorithm successfully samples a configuration within $\text{polylog}(n)$ rounds in the LOCAL model with high probability. In particular, this algorithm samples random

q -coloring of triangle-free graphs within $O(\log^3 n)$ rounds in the LOCAL model as long as $q \geq (\alpha^* + \delta)\Delta$. This non-MCMC sampling algorithm abuses the power of the LOCAL model by assuming unlimited message-size and local computations.

It is a well known phenomenon that sampling may become computationally intractable when the model exhibits the non-uniqueness phase-transition property, e.g. independent sets in graphs of maximum degree bounded by a $\Delta \geq 6$ [27, 28,55,56]. For the same class of distributions, we show the following unconditional $\Omega(\text{diam})$ lower bound for sampling in the LOCAL model.

Theorem 3 *For $\Delta \geq 6$, there exist infinitely many graphs $G(V, E)$ with maximum degree Δ and diameter $\text{diam}(G) = |V|^{\Omega(1)}$ such that any algorithm that samples uniform independent set in G within sufficiently small constant total variation distance ϵ requires $\Omega(\text{diam}(G))$ rounds of communications, even assuming the vertices $v \in V$ to be aware of G .*

The lower bound is proved by a now fairly well-understood reduction from maximum cut to sampling independent sets when $\Delta \geq 6$ [28,55,56]. Specifically, we show that when $\Delta \geq 6$ there are infinitely many graphs $G(V, E)$ such that if one can sample a nearly uniform independent set in $G(V, E)$, then one can also sample an almost uniform maximum cut in an even cycle of size $|V|^{\Omega(1)}$, which is necessarily a global task because of the long-range correlation.

Theorem 3 strongly separates sampling from labeling problems for distributed computing:

- In the LOCAL model it is trivial to construct an independent set (because \emptyset is an independent set). In contrast, Theorem 3 says that sampling a uniform independent set is very much a global task for graphs with maximum degree $\Delta \geq 6$.
- In the LOCAL model any labeling problem would be trivial once the network structure G is known to each vertex. In contrast, the sampling lower bound in Theorem 3 still holds even when each vertex is aware of G . Unlike labeling whose hardness is due to the locality of information, for sampling the hardness is solely due to the locality of randomness.
- A breakthrough of Ghaffari et al. [30] shows that any labeling problem that can be solved sequentially with local information admits a $O(\text{polylog}(n))$ -round randomized protocol in the LOCAL model. In contrast, for sampling we have an $\Omega(\text{diam})$ randomized lower bound for graphs with $n^{\Omega(1)}$ diameter.

1.2 Related work

The topic of sequential MCMC (Markov chain Monte Carlo) sampling is extensively studied. The study of sampling proper q -colorings was initiated by the seminal works of Jerrum [37] and independently of Salas and Sokal [53]. So far the best rapid mixing condition for general bounded-degree graphs is $q \geq \frac{11}{6} \Delta$ due to Vigoda [59]. See [26] for an excellent survey.

The chromatic-scheduler-based parallelization of Glauber dynamics was studied in [32]. This parallel chain is in fact a special case of systematic scan for Glauber dynamics [18, 19, 35], in which the variables are updated according to a fixed order.

Empirical studies showed that sometimes an *ad hoc* “Hogwild!” parallelization of sequential sampler might work well in practice [51] and the mixing results assuming bounded asynchrony were given in [14, 38].

A sampling algorithm based on the Lovász local lemma is given in [33]. When sampling from the hardcore model with $\lambda < \frac{1}{2\sqrt{e}\Delta-1}$ on a graph of maximum degree Δ , this sampling algorithm can be implemented in the LOCAL model which runs in $O(\log n)$ rounds.

A problem related to the local sampling is the *finitary coloring* [36], in which a random feasible solution is sampled according to an unconstrained distribution as long as the distribution is over feasible solutions, rather than a specific distribution such as the Gibbs distribution. Therefore, the nature of this problem is still labeling rather than sampling.

Our algorithms are Markov chains which randomly walk over the solution space. A related notion is the *distributed random walks* [13], which walk over the network.

Our *LocalMetropolis* algorithm should be distinguished from the parallel Metropolis-Hastings algorithm [9] or the parallel tempering [58], in which the sampling algorithms makes N proposals or runs N copies of the system in parallel for a suitably large N , in order to improve the dynamic properties of the Monte Carlo simulation.

Organization of the paper The models and preliminaries are introduced in Sect. 2. The *LubyGlauber* algorithm is introduced in Sect. 3. The *LocalMetropolis* algorithm is introduced in Sect. 4. And the lower bounds are proved in Sect. 5.

2 Models and preliminaries

2.1 The LOCAL model

We assume Linial’s LOCAL model [49, 52] for distributed computation, which is as described in Sect. 1. We further allow each node in the network $G(V, E)$ to be aware of

upper bounds of Δ and $\log n$, where $n = |V|$ is the number of nodes. This information is accessed only because the running time of the Monte Carlo algorithms may depend on them.

2.2 Markov random field and local CSP

The *Markov random field (MRF)*, or *spin system*, is a well studied stochastic model in probability theory and statistical physics. Given a graph $G(V, E)$ and a set of *spin states* $[q] = \{1, 2, \dots, q\}$ for a finite $q \geq 2$, a *configuration* $\sigma \in [q]^V$ assigns each vertex one of the q spin states. For each edge $e \in E$ there is a non-negative $q \times q$ symmetric matrix $A_e \in \mathbb{R}_{\geq 0}^{q \times q}$ associated with e , called the *edge activity*; and for each vertex $v \in V$ there is a non-negative q -dimensional vector $b_v \in \mathbb{R}_{\geq 0}^q$ associated with v , called the *vertex activity*. Then each configuration $\sigma \in [q]^V$ is assigned a weight $w(\sigma)$ which is as defined in (1).

This gives rise to a natural probability distribution μ , called the *Gibbs distribution*, over all configurations in the sample space $\Omega = [q]^V$ proportional to their weights, such that $\mu(\sigma) = w(\sigma)/Z$ for each $\sigma \in \Omega$, where $Z = \sum_{\sigma \in \Omega} w(\sigma)$ is the normalizing factor. A configuration $\sigma \in \Omega$ is *feasible* if $\mu(\sigma) > 0$.

Several natural joint distributions can be expressed as MRFs:

- *Independent sets/vertex covers*: When $q = 2$, all $A_e = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ and all $b_v = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, each feasible configuration corresponds to an independent set (or vertex cover, if the other spin state indicates the set) in G , and the Gibbs distribution μ is the uniform distribution over independent sets (or vertex covers) in G . When $b_v = \begin{bmatrix} 1 \\ \lambda \end{bmatrix}$ for some parameter $\lambda > 0$, this is the *hardcore model* from statistical physics.
- *Colorings and list colorings*: When every A_e has $A_e(i, i) = 0$ and $A_e(i, j) = 1$ if $i \neq j$, and every b_v is the all-1 vector, the Gibbs distribution μ becomes the uniform distribution over proper q -colorings of graph G . For *list colorings*, each vertex $v \in V$ can only use the colors from its list $L_v \subseteq [q]$ of available colors. Then we can let each b_v be the indicator vector for the list L_v and A_e ’s are the same as for proper q -colorings, so that the Gibbs distribution is the uniform distribution over proper list colorings.
- *Physical model*: The proper q -coloring is a special case of the *Potts model* in statistical physics, in which each A_e has $A_e(i, i) = \beta$ for some parameter $\beta > 0$ and $A_e(i, j) = 1$ if $i \neq j$. When further $q = 2$, the model becomes the *Ising model*.

The model of MRF can be further generalized to allow multivariate asymmetric constraints, by which gives us the *weighted CSPs*, also known as the *factor graphs*. In this model, we have a collection \mathcal{C} of *constraints* $c = (f_c, S_c)$ where each $f_c : [q]^{|S_c|} \rightarrow \mathbb{R}_{\geq 0}$ is a *constraint function* with *scope* $S_c \subseteq V$. Each configuration $\sigma \in [q]^V$ is assigned a weight:

$$w(\sigma) = \prod_{c=(f_c, S_c) \in \mathcal{C}} f_c(\sigma|_{S_c}),$$

where $\sigma|_{S_c}$ represents the restriction of σ on S_c . And the Gibbs distribution μ over all configurations in $\Omega = [q]^V$ is defined in the same way proportional to the weights. In particular, when f_c 's are Boolean-valued functions, the Gibbs distribution μ is the uniform distribution over CSP solutions.

A constraint $c = (f_c, S_c)$ is said to be *local* with respect to network G if the diameter of the scope S_c in network G is bounded by a constant. Local CSPs are expressive, for example:

- *Dominating sets*: They can be expressed by having a ‘‘cover’’ constraint on each inclusive neighborhood $\Gamma^+(v)$ which constrains that at least one vertex from $\Gamma^+(v)$ is chosen.
- *Maximal independent sets (MISs)*: An MIS is a dominating independent set.

Clearly, the MRF is a special class of weighted local CSPs, defined by unary and binary symmetric local constraints with respect to G .

2.3 Local sampling

The local sampling problem is defined as follows. Let $G(V, E)$ be a network. Given an MRF defined on G (or more generally a weighted CSP that is local with respect to G), where the specifications of the local constraints are given as private inputs to the involved processors, for any $\epsilon > 0$ upon termination each processor $v \in V$ outputs a random variable X_v such that the total variation distance between the distribution ν of the random vector $X = (X_v)_{v \in V}$ and the Gibbs distribution μ is bounded as $d_{TV}(\mu, \nu) \leq \epsilon$, where the total variation distance between two distributions μ, ν over $\Omega = [q]^V$ is defined as

$$d_{TV}(\mu, \nu) = \sum_{\sigma \in \Omega} \frac{1}{2} |\mu(\sigma) - \nu(\sigma)| = \max_{A \subseteq \Omega} |\mu(A) - \nu(A)|.$$

2.4 Mixing rate

Our algorithms are given as Markov chains. Given an *irreducible* and *aperiodic* Markov chain $X^{(0)}, X^{(1)}, \dots \in \Omega$, for

any $\sigma \in \Omega$ let $\pi_\sigma^{(t)}$ denote the distribution of $X^{(t)}$ conditioning on that $X^{(0)} = \sigma$. For $\epsilon > 0$ the *mixing rate* $\tau(\epsilon)$ is defined as

$$\tau(\epsilon) = \max_{\sigma \in \Omega} \min \left\{ t : d_{TV}(\pi_\sigma^{(t)}, \pi) \leq \epsilon \right\},$$

where π is the *stationary distribution* for the chain. For formal definitions of these notions for Markov chain, we refer to a standard textbook of the subject [43]. Informally, irreducibility and aperiodicity guarantees that $X^{(t)}$ converges to the unique stationary distribution π as $t \rightarrow \infty$, and the mixing rate $\tau(\epsilon)$ tells us how fast it converges.

Notations Given a graph $G(V, E)$, we denote by $d_v = \deg(v)$ the degree of v in G , $\Delta = \Delta_G$ the maximum degree of G , $\text{diam} = \text{diam}(G)$ the diameter of G , and $\text{dist}(u, v) = \text{dist}_G(u, v)$ the shortest path distance between vertices u and v in G .

We also denote by $\Gamma(v) = \{u \mid uv \in E\}$ the neighborhood of v , and $\Gamma^+(v) = \Gamma(v) \cup \{v\}$ the inclusive neighborhood. Finally we write $B_r(v) = \{u \mid \text{dist}(u, v) \leq r\}$ for the r -ball centered at v .

3 The LubyGlauber algorithm

In this section, we analyze a generic scheme for parallelizing Glauber dynamics, a classic sequential Markov chain for sampling from Gibbs distributions.

We assume a Markov random field (MRF) defined on the network $G(V, E)$, with edge activities $A = \{A_e\}_{e \in E}$ and vertex activities $\mathbf{b} = \{b_v\}_{v \in V}$, which specifies a Gibbs distribution μ over $\Omega = [q]^V$. The *single-site heat-bath Glauber dynamics*, or simply the *Glauber dynamics*, is a well known Markov chain for sampling from the Gibbs distribution μ . Starting from an arbitrary initial configuration $X \in [q]^V$, at each step the chain does the followings:

- sample a vertex $v \in V$ uniformly at random;
- resample the value of X_v according to the marginal distribution induced by μ at vertex v conditioning on the current spin states of v 's neighborhood.

It is well known (see [43]) that the Glauber dynamics is a reversible Markov chain whose stationary distribution is the Gibbs distribution μ .

Formally, supposed that $\sigma \in [q]^V$ is sampled from μ , for any $v \in V, S \subseteq V$ and $\tau_S \in [q]^S$, the *marginal distribution* at vertex v conditioning on τ_S , denoted as $\mu_v(\cdot \mid \tau_S)$, is defined as

$$\forall c \in [q], \quad \mu_v(c \mid \tau_S) = \Pr[\sigma_v = c \mid \sigma_S = \tau_S].$$

In the Glauber dynamics, X_v is resampled according to the marginal distribution $\mu_v(\cdot \mid X_{\Gamma(v)})$. Here $X_{\Gamma(v)}$ represents the current spin states of v 's neighborhood $\Gamma(v)$. For Markov random field, this marginal distribution can be computed as

$$\forall c \in [q], \mu_v(c \mid X_{\Gamma(v)}) = \frac{b_v(c) \prod_{u \in \Gamma(v)} A_{uv}(c, X_u)}{\sum_{a \in [q]} b_v(a) \prod_{u \in \Gamma(v)} A_{uv}(a, X_u)}. \tag{2}$$

For example, when the MRF is the proper q -coloring, this is just the uniform distribution over available colors in $[q]$ which are not used by v 's neighbors. For the Glauber dynamics to work, it is common to assume that the sum $\sum_{a \in [q]} b_v(a) \prod_{u \in \Gamma(v)} A_{uv}(a, X_u)$ is always positive, so that the marginal distributions are well-defined.¹

A generic scheme for parallelizing the Glauber dynamics is that at each step, instead of updating one vertex, the chain updates a group of “non-interfering” vertices in parallel, as follows:

- independently sample a random independent set I in G ;
- for each $v \in I$, resample X_v in parallel according to the marginal distribution $\mu_v(\cdot \mid X_{\Gamma(v)})$.

This can be seen as a relaxation of the chromatic-based scheduler [32] and systematic scans [19].

A convenient way for generating a random independent set in a distributed fashion is the “Luby step” in Luby’s algorithm for distributed MIS [1,46]: each vertex samples a uniform and independent ID from the interval $[0, 1]$ (which can be discretized with $O(\log n)$ bits) and the vertices v who are locally maximal among the inclusive neighborhood $\Gamma^+(v)$ are selected into the independent set I .

The resulting algorithm is called *LubyGlauber*, whose pseudocode is given in Algorithm 1.

Algorithm 1: Pseudocode for vertex $v \in V$ in *LubyGlauber* algorithm

Input: Vertex $v \in V$ receives $\{A_{uv}\}_{u \in \Gamma(v)}$ and b_v as input.
 1 initialize X_v to an arbitrary value in $[q]$;
 2 **for** $t = 1$ **through** T **do**
 3 sample a real $\beta_v \in [0, 1]$ uniformly and independently;
 4 **if** $\beta_v > \max\{\beta_u \mid u \in \Gamma(v)\}$ **then**
 5 resample X_v according to marginal distribution $\mu_v(\cdot \mid X_{\Gamma(v)})$;
 6 **return** X_v ;

¹ This property holds automatically for feasible configurations X with $\mu(X) > 0$, and is only needed when the Glauber dynamics is allowed to start from an infeasible configuration. For specific MRF, such as proper q -coloring, this property is guaranteed by the “uniqueness condition” $q \geq \Delta + 1$.

According to the definition of marginal distribution (2), resampling X_v can be done locally by exchanging neighbors’ current spin states. After T iterations, where T is a threshold determined for specific Markov random field, the algorithm terminates and outputs the current $X = (X_v)_{v \in V}$.

Remark 1 The *LubyGlauber* algorithm can be easily extended to sample from weighted CSPs defined by local constraints $c = (f_c, S_c) \in \mathcal{C}$, by simply overriding the definition of neighborhood as $\Gamma(v) = \{u \neq v \mid \exists c \in \mathcal{C}, \{u, v\} \subseteq S_c\}$, thus $\Gamma(v)$ is the neighborhood of v in the hypergraph where S_c 's are the hyperedges and now I is the strongly independent set of this hypergraph.

3.1 Mixing of *LubyGlauber*

Let μ_{LG} denote the distribution of X returned by the algorithm upon termination. As in the case of single-site Glauber dynamics, we assume that the marginal distribution (2) is always well-defined, and the single-site Glauber dynamics is irreducible among all feasible configurations. The following proposition is easy to obtain.

Proposition 1 *The Markov chain LubyGlauber is reversible and has stationary distribution μ . Furthermore, under the above assumption, $d_{TV}(\mu_{LG}, \mu)$ converges to 0 as $T \rightarrow \infty$.*

Proof We prove this for a more general family of Markov chains, where the “Luby step” is replaced by an arbitrary way of independently sampling a random independent set I , as long as $\Pr[v \in I] > 0$ for every vertex $v \in V$.

Let $\Omega = [q]^V$ and $P \in \mathbb{R}_{\geq 0}^{|\Omega| \times |\Omega|}$ denote the transition matrix for the *LubyGlauber* chain. We first show that the chain is reversible and μ is stationary. Specifically, this means to verify the *detailed balance equation*:

$$\mu(X)P(X, Y) = \mu(Y)P(Y, X),$$

for all configurations $X, Y \in \Omega = [q]^V$.

If both X and Y are infeasible, then $\mu(X) = \mu(Y) = 0$ and the detailed balance equation holds trivially. If X is feasible and Y is not then $\mu(Y) = 0$ and meanwhile since the chain never moves from a feasible configuration to an infeasible one, we have $P(X, Y) = 0$ so the detailed balance equation is also satisfied.

It remains to verify the detailed balance equation when both X and Y are feasible. Let $D = \{v \in V \mid X_v \neq Y_v\}$ be the set of disagreeing vertices. If D is not an independent set, then $P(X, Y) = P(Y, X) = 0$ and the detailed balance equation holds. Suppose that D is an independent set. For any independent set $I \supseteq D$, we denote by $\Pr[X \rightarrow Y \mid I]$ the probability that within an iteration the chain moves from X to Y conditioning on I being the independent set sampled in the first step. Therefore,

$$\begin{aligned} \frac{\Pr[X \rightarrow Y \mid I]}{\Pr[Y \rightarrow X \mid I]} &= \frac{\prod_{v \in D} b_v(Y_v) \prod_{u \in \Gamma(v)} A_{uv}(Y_u, Y_v)}{\prod_{v \in D} b_v(X_v) \prod_{u \in \Gamma(v)} A_{uv}(X_u, X_v)} \\ &= \frac{\mu(Y)}{\mu(X)}. \end{aligned}$$

By the law of total probability,

$$\begin{aligned} \frac{P(X, Y)}{P(Y, X)} &= \frac{\sum_{I \supseteq D} \Pr(I) \Pr[X \rightarrow Y \mid I]}{\sum_{I \supseteq D} \Pr(I) \Pr[Y \rightarrow X \mid I]} \\ &= \frac{\prod_{v \in D} b_v(Y_v) \prod_{u \in \Gamma(v)} A_{uv}(Y_u, Y_v)}{\prod_{v \in D} b_v(X_v) \prod_{u \in \Gamma(v)} A_{uv}(X_u, X_v)} \\ &= \frac{\mu(Y)}{\mu(X)}. \end{aligned}$$

Thus, the chain is reversible with respect to the Gibbs distribution μ .

Next, observe that the chain will never move from a feasible configuration to an infeasible one. Moreover, due to the assumption that the marginal distribution (2) is always well-defined, once a vertex v has been resampled, it will satisfy all local constraints. Therefore, the chain will be feasible once every vertex has been resampled. Since every vertex v has positive probability $\Pr[v \in I]$ to be resampled, the chain is absorbing to feasible configurations.

It is easy to observe that every feasible configuration is aperiodic, since it has self-loop transition, i.e. $P(X, X) > 0$ for all feasible X . And any move $X \rightarrow Y$ between feasible configurations $X, Y \in \Omega$ in the single-site Glauber dynamics with vertex v being updated, can be simulated by a move in the *LubyGlauber* chain by first sampling an independent set $I \ni v$ (which is always possible since $\Pr[v \in I] > 0$) and then updating v according to $X \rightarrow Y$ and meanwhile keeping all $v \in I \setminus \{v\}$ unchanged (which is always possible for feasible X). Provided the irreducibility of the single-site Glauber dynamics among all feasible configurations, the *LubyGlauber* chain is also irreducible among all feasible configurations. Combining with the absorption towards feasible configurations and their aperiodicity, due to the Markov chain convergence theorem [43], the total variation distance $d_{TV}(\mu_{LG}, \mu)$ converges to 0 as $T \rightarrow \infty$. \square

We then apply a standard coupling argument from [18, 35] to analyze the mixing rate of the *LubyGlauber* chain. The following notions are essential to the mixing of Glauber dynamics.

Definition 1 (influence matrix) For $v \in V$ and $\sigma \in [q]^V$, we write $\mu_v^\sigma = \mu_v(\cdot \mid \sigma_{\Gamma(v)})$ for the marginal distribution of the value of v , for configurations sampled from μ conditioning on agreeing with σ at all neighbors of v . For vertices $i, j \in V$, the *influence* of j on i is defined as

$$\rho_{i,j} := \max_{(\sigma, \tau) \in S_j} d_{TV}(\mu_i^\sigma, \mu_i^\tau),$$

where S_j denotes the set of all pairs of feasible configurations $\sigma, \tau \in [q]^V$ such that σ and τ agree on all vertices except j . Let $R = (\rho_{i,j})_{i,j \in V}$ be the $n \times n$ influence matrix.

Definition 2 (Dobrushin’s condition) Let α be the total influence on a vertex, defined by

$$\alpha := \max_{i \in V} \sum_{j \in V} \rho_{i,j}.$$

We say that the Dobrushin’s condition is satisfied if $\alpha < 1$.

It is a fundamental result that the Dobrushin’s condition is sufficient for the rapid mixing of Glauber dynamics [16,35, 53], with a mixing rate of $\tau(\epsilon) = O\left(\frac{n}{1-\alpha} \log\left(\frac{n}{\epsilon}\right)\right)$. Here we show that the *LubyGlauber* chain is essentially a parallel speed up of the Glauber dynamics by a factor of $\Theta\left(\frac{n}{\Delta}\right)$.

Theorem 4 Under the same assumption as Proposition 1, if the total influence $\alpha < 1$, then the mixing rate of the *LubyGlauber* chain is $\tau(\epsilon) = O\left(\frac{\Delta}{1-\alpha} \log\left(\frac{n}{\epsilon}\right)\right)$.

Consequently, for any $\epsilon > 0$ the *LubyGlauber* algorithm can terminate within $O\left(\frac{\Delta}{1-\alpha} \log\left(\frac{n}{\epsilon}\right)\right)$ rounds in the LOCAL model and return an $X \in [q]^V$ whose distribution μ_{LG} is ϵ -close to the Gibbs distribution μ in total variation distance.

Remark 2 In fact, Proposition 1 and Theorem 4 hold for a more general family of Markov chains, where the “Luby step” could be any subroutine which independently generates a random independent set I , as long as every vertex has positive probability to be selected into I . In general, the mixing rate in Theorem 4 is in fact $\tau(\epsilon) = O\left(\frac{1}{(1-\alpha)^\gamma} \log\left(\frac{n}{\epsilon}\right)\right)$ where γ is a lower bound for the probability $\Pr[v \in I]$ for all $v \in V$.

The following lemma is crucial for relating the mixing rate to the influence matrix. The lemma has been proved in various places [14,18,35].

Lemma 1 Let X and Y be two random variables that take values over the feasible configurations in $\Omega = [q]^V$, then for any $i \in V$,

$$\mathbf{E}_{(X,Y)} \left[d_{TV}(\mu_i^X, \mu_i^Y) \right] \leq \sum_{k \in V} \rho_{i,k} \Pr[X_k \neq Y_k].$$

Proof We enumerate V as $V = \{1, 2, \dots, n\}$. For $0 \leq k \leq n$, define $Z^{(k)}$ as that for each $j \in V$, $Z_j^{(k)} = X_j$ if $j > k$ and $Z_j^{(k)} = Y_j$ if $j \leq k$. In particular, $Z^{(0)} = X$ and $Z^{(n)} = Y$. Now, by triangle inequality,

$$\begin{aligned} d_{TV}(\mu_i^X, \mu_i^Y) &= d_{TV}(\mu_i^{Z^{(0)}}, \mu_i^{Z^{(n)}}) \\ &\leq \sum_{k=1}^n d_{TV}(\mu_i^{Z^{(k-1)}}, \mu_i^{Z^{(k)}}). \end{aligned}$$

Next, we note that $Z^{(k-1)} = Z^{(k)}$ if and only if $X_k = Y_k$. Therefore,

$$d_{TV}(\mu_i^X, \mu_i^Y) \leq \sum_{k=1}^n \mathbf{1}\{X_k \neq Y_k\} d_{TV}(\mu_i^{Z^{(k-1)}}, \mu_i^{Z^{(k)}}).$$

Since $Z^{(k-1)}$ and $Z^{(k)}$ can only differ at vertex k , it follows that $(Z^{(k-1)}, Z^{(k)}) \in S_k$, and hence,

$$\begin{aligned} d_{TV}(\mu_i^X, \mu_i^Y) &\leq \sum_{k=1}^n \mathbf{1}\{X_k \neq Y_k\} \max_{(\sigma, \tau) \in S_k} d_{TV}(\mu_i^\sigma, \mu_i^\tau) \\ &= \sum_{k=1}^n \rho_{i,k} \mathbf{1}\{X_k \neq Y_k\}. \end{aligned}$$

By linearity of expectation,

$$\mathbf{E}_{(X,Y)} \left[d_{TV}(\mu_i^X, \mu_i^Y) \right] \leq \sum_{k \in V} \rho_{i,k} \Pr[X_k \neq Y_k].$$

□

Proof of Theorem 4: We are actually going to prove a stronger result: Denoted by I the random independent set on which the resampling is executed, we write $\gamma_v = \Pr[v \in I]$ for each $v \in V$, and assume that for all $v \in V$, $\gamma_v \geq \gamma$ for some $\gamma > 0$. Clearly, when I is generated by the ‘‘Luby step’’, this holds for $\gamma = \frac{1}{\Delta+1}$. We are going to prove that $\tau(\epsilon) = O\left(\frac{1}{(1-\alpha)\gamma} \log\left(\frac{n}{\epsilon}\right)\right)$.

The proof follows the framework of Hayes [35]. We construct a coupling of the Markov chain $(X^{(t)}, Y^{(t)})$ such that the transition rules for $X^{(t)} \rightarrow X^{(t+1)}$ and $Y^{(t)} \rightarrow Y^{(t+1)}$ are the same as the *LubyGlauber* chain. If $\Pr[X^{(T)} \neq Y^{(T)} \mid X^{(0)} = \sigma \wedge Y^{(0)} = \tau] \leq \epsilon$ for any initial configurations $\sigma, \tau \in \Omega$, then by the coupling lemma for Markov chain [43], we have the mixing rate $\tau(\epsilon) \leq T$.

The coupling we are going to use is the maximal one-step coupling of the *LubyGlauber* chain, which for every vertex $i \in V$ achieves that

$$\Pr[X_i^{(t+1)} \neq Y_i^{(t+1)} \mid X^{(t)}, Y^{(t)}] = d_{TV}(\mu_i^{X^{(t)}}, \mu_i^{Y^{(t)}}),$$

where $\mu_i^{X^{(t)}}$ and $\mu_i^{Y^{(t)}}$ are the marginal distributions as defined in Definition 1. The existence of such coupling is guaranteed by the coupling lemma.

Arbitrarily fix $\sigma, \tau \in \Omega = [q]^V$. For $t \geq 0$, define $(X^{(t)}, Y^{(t)}) \in \Omega^2$ by iterating a maximal one-step coupling of the *LubyGlauber* chain, starting from initial condition $X^{(0)} = \sigma, Y^{(0)} = \tau$. Due to the well-defined-ness of marginal distribution (2), we know that once all vertices have been resampled, the configuration will be feasible and will remain to be feasible in future.

Let T_1 be a positive integer and \mathcal{F} denote the event all vertices have been resampled in chain X and Y in the first T_1 steps. By union bound, we have

$$\Pr[\neg \mathcal{F}] \leq 2 \sum_{v \in V} (1 - \gamma_v)^{T_1} \leq 2n(1 - \gamma)^{T_1}, \tag{3}$$

Next, we assume that $X^{(t)}, Y^{(t)}$ are both feasible for $t \geq T_1$. We define the vector $\mathbf{p}^{(t)} \in [0, 1]^V$ as

$$\forall j \in V, \quad p_j^{(t)} := \Pr[X_j^{(t)} \neq Y_j^{(t)}].$$

By the definition of the *LubyGlauber* chain, it holds for every $j \in V$ that

$$p_j^{(t+1)} = (1 - \gamma_j) p_j^{(t)} + \gamma_j \cdot \Pr[X_j^{(t+1)} \neq Y_j^{(t+1)} \mid j \in I]. \tag{4}$$

By the definition of maximal one-step coupling and Lemma 1, for $t \geq T_1$, for any $i \in V$,

$$\begin{aligned} &\Pr[X_i^{(t+1)} \neq Y_i^{(t+1)} \mid i \in I] \\ &= \sum_{\substack{\sigma, \tau \in \Omega \\ \mu(\sigma), \mu(\tau) > 0}} \left(\Pr[X_i^{(t+1)} \neq Y_i^{(t+1)} \mid X^{(t)} = \sigma, Y^{(t)} = \tau] \right. \\ &\quad \left. \cdot \Pr[X^{(t)} = \sigma \wedge Y^{(t)} = \tau] \right) \\ &= \sum_{\substack{\sigma, \tau \in \Omega \\ \mu(\sigma), \mu(\tau) > 0}} d_{TV}(\mu_i^\sigma, \mu_i^\tau) \cdot \Pr[X^{(t)} = \sigma \wedge Y^{(t)} = \tau] \\ &= \mathbf{E} \left[d_{TV}(\mu_i^{X^{(t)}}, \mu_i^{Y^{(t)}}) \right] \\ &\leq \sum_{k \in V} \rho_{i,k} \cdot \Pr[X_k^{(t)} \neq Y_k^{(t)}]. \end{aligned}$$

Combined with equality (4), for $t \geq T_1$ we have

$$\mathbf{p}^{(t+1)} \leq M \mathbf{p}^{(t)},$$

where matrix $M = (J - \Gamma)J + \Gamma R$, where Γ is the $n \times n$ diagonal matrix with $\Gamma_{i,i} = \gamma_i$; J is the $n \times n$ identity matrix; and $R = (\rho_{ij})$ is the influence matrix. The ∞ -norm of M is bounded as

$$\begin{aligned} \|M\|_\infty &= \max_{i \in V} \sum_{j \in V} |M_{i,j}| \\ &\leq \max_{i \in V} \{1 - (1 - \alpha)\gamma_i\} \\ &\leq 1 - (1 - \alpha)\gamma. \end{aligned}$$

Let $T = T_1 + T_2$. By induction, we obtain the component-wise inequality

$$\mathbf{p}^{(T)} \leq M^{T_2} \mathbf{p}^{(T_1)}.$$

Conditioning on that $X^{(T_1)}$ and $Y^{(T_1)}$ are both feasible, we have

$$\begin{aligned} \Pr \left[X^{(T)} \neq Y^{(T)} \right] &\leq \|\mathbf{p}^{(T)}\|_1 \quad \text{by union bound} \\ &\leq n \|\mathbf{p}^{(T)}\|_\infty \quad \text{by Hölder's inequality} \\ &\leq n \|M^{T_2} \mathbf{p}^{(T_1)}\|_\infty \\ &\leq n \|M\|_\infty^{T_2} \|\mathbf{p}^{(T_1)}\|_\infty \\ &\leq n (1 - (1 - \alpha)\gamma)^{T_2} \end{aligned} \tag{5}$$

For any ϵ , we choose $T_1 = \left\lceil \frac{1}{\gamma} \ln \left(\frac{4n}{\epsilon} \right) \right\rceil$ and $T_2 = \left\lceil \frac{1}{(1-\alpha)\gamma} \ln \left(\frac{2n}{\epsilon} \right) \right\rceil$. Then $T = T_1 + T_2 = O \left(\frac{1}{(1-\alpha)\gamma} \log \left(\frac{n}{\epsilon} \right) \right)$. Combining (3) and (5), conditioning on $X^{(0)} = \sigma \wedge Y^{(0)} = \tau$ for arbitrary $\sigma, \tau \in \Omega$, we have

$$\begin{aligned} \Pr \left[X^{(T)} \neq Y^{(T)} \right] &\leq \Pr[\neg \mathcal{F}] + \Pr \left[X^{(T)} \neq Y^{(T)} \mid \mathcal{F} \right] \\ &\leq 2n(1 - \gamma)^{T_1} + n(1 - (1 - \alpha)\gamma)^{T_2} \\ &\leq \epsilon. \end{aligned}$$

This implies that

$$\tau(\epsilon) = O \left(\frac{1}{(1 - \alpha)\gamma} \log \left(\frac{n}{\epsilon} \right) \right).$$

In particular, if the random independent set I is generated by the ‘‘Luby step’’, we have $\gamma = \frac{1}{\Delta+1}$, therefore for the LubyGlauber chain

$$\tau(\epsilon) = O \left(\frac{\Delta}{1 - \alpha} \log \left(\frac{n}{\epsilon} \right) \right).$$

□

3.2 Application of LubyGlauber for sampling graph colorings

For uniformly distributed proper q -coloring of graph G , it is well known that the Dobrushin’s condition is satisfied when $q \geq 2\Delta + 1$ where Δ is the maximum degree of graph G .

We consider a more generalized problem, the list colorings, where each vertex $v \in V$ maintains a list $L_v \subseteq [q]$ of colors that it can use. The proper q -coloring is a special case of list coloring when everyone’s list is precisely $[q]$. For each vertex $v \in V$, we denote by $q_v = |L_v|$ the size of v ’s list,

and $d_v = \deg(v)$ the degree of v . It is easy to verify that the total influence α is now bounded as

$$\alpha = \max_{i \in V} \sum_{j \in V} \rho_{i,j} = \max_{v \in V} \left\{ \frac{d_v}{q_v - d_v} \right\}.$$

Applying Theorem 4, we have the following corollary, which also implies Theorem 1.

Corollary 1 *If there is an arbitrary constant $\delta > 0$ such that $q_v \geq (2 + \delta)d_v$ for every vertex v , then the mixing rate of the LubyGlauber chain for sampling list coloring is $\tau(\epsilon) = O \left(\Delta \log \left(\frac{n}{\epsilon} \right) \right)$.*

4 The LocalMetropolis algorithm

In this section, we give an algorithm that may fully parallelize the sequential process under suitable mixing conditions, even on graphs with unbounded degree. The algorithm is inspired by the famous Metropolis-Hastings algorithm for MCMC, in which a random choice is *proposed* and then *filtered* to enforce the target stationary distribution. Our algorithm, called the *LocalMetropolis* algorithm, makes each vertex propose independently, and localizes the work of filtering to each edge.

We are given a Markov random field (MRF) defined on the network $G(V, E)$, with edge activities $\mathbf{A} = \{A_e\}_{e \in E}$ and vertex activities $\mathbf{b} = \{b_v\}_{v \in V}$, whose Gibbs distribution is μ . Starting from an arbitrary configuration $X \in [q]^V$, in each iteration, the *LocalMetropolis* chain does the followings:

- **Propose:** Each vertex $v \in V$ independently proposes a spin state $\sigma_v \in [q]$ with probability proportional to $b_v(\sigma_v)$.
- **Local filter:** Each edge $e \in E$ flips a biased coin independently, with the probability of HEADS being

$$\tilde{A}_e(\sigma_u, \sigma_v) \tilde{A}_e(X_u, \sigma_v) \tilde{A}_e(\sigma_u, X_v),$$

where \tilde{A}_e is the matrix obtained by normalizing A_e as $\tilde{A}_e = A_e / \max_{i,j} A_e(i, j)$. We say that the edge *passes the check* if the outcome of coin flipping is HEADS.

Then for each vertex $v \in V$, if all edges incident with v passed their checks, v accepts the proposal and updates the value as $X_v = \sigma_v$, otherwise v leaves X_v unchanged.

After T iterations, where T is a threshold determined for specific Markov random field, the algorithm terminates and outputs the current $\mathbf{X} = (X_v)_{v \in V}$. The pseudocode for the *LocalMetropolis* algorithm is given in Algorithm 2.

We remark that in each iteration, for each edge $e = uv$, the two endpoints u and v access the same random coin to determine whether e passes the check in this iteration.

Algorithm 2: Pseudocode for the *LocalMetropolis* algorithm

Input: Each vertex $v \in V$ receives $\{A_{uv}\}_{u \in \Gamma(v)}$ and b_v as input.

- 1 each $v \in V$ initializes X_v to an arbitrary value in $[q]$;
- 2 **for** $t = 1$ **through** T **do**
- 3 **foreach** $v \in V$ **do**
- 4 propose a random $\sigma_v \in [q]$ with probability
 $b_v(\sigma_v) / \sum_{c \in [q]} b_v(c)$;
- 5 **foreach** $e = (u, v) \in E$ **do**
- 6 pass the check independently with probability
 $\frac{A_e(\sigma_u, \sigma_v) A_e(X_u, \sigma_v) A_e(\sigma_u, X_v)}{(\max_{i, j \in [q]} A_e(i, j))^2}$;
- 7 **foreach** $v \in V$ **do**
- 8 **if** all edges e incident with v pass the checks **then**
- 9 $X_v \leftarrow \sigma_v$;
- 10 each $v \in V$ **returns** X_v ;

Remark 3 The *LocalMetropolis* algorithm can be naturally extended to sample from weighted CSPs. The local filtering now occurs on each local constraint, such that a k -ary constraint $c = (f_c, S_c) \in \mathcal{C}$ passes the check with the probability which is a product of $2^k - 1$ normalized factors $\tilde{f}_c(\tau)$ for the $\tau \in [q]^{S_c}$ obtained from $2^k - 1$ ways of mixing σ_{S_c} with X_{S_c} except the X_{S_c} itself.

4.1 Mixing of LocalMetropolis

Let μ_{LM} denote the distribution of $X = (X_v)_{v \in V}$ returned by the *LocalMetropolis* algorithm after T iterations.

We need to ensure the chain is well behaved even when starting from infeasible configurations. Now we make the following assumption: for all $X \in [q]^V$ and $v \in V$,

$$\sum_{i \in [q]} b_v(i) \prod_{u \in \Gamma(v)} A_{uv}(i, X_u) \sum_{j \in [q]} b_u(j) A_{uv}(X_v, j) A_{uv}(i, j) > 0, \tag{6}$$

which is slightly stronger than the assumption made for the Glauber dynamics. As in the case of Glauber dynamics, the property is needed only when the chain is allowed to start from an infeasible configuration $X \in [q]^V$ with $\mu(X) = 0$. For specific MRF, such as graph colorings, the condition (6) is satisfied as long as $q \geq \Delta + 1$ and $q \geq 3$. As before, we further assume that the single-site Markov chain² is irreducible among feasible configurations.

Theorem 5 *The Markov chain LocalMetropolis is reversible and has stationary distribution μ . Furthermore, under above assumptions, $d_{TV}(\mu_{LM}, \mu)$ converges to 0 as $T \rightarrow \infty$.*

² For the MRFs, since the single-site Glauber dynamics has the same connectivity structure as the natural single-site version of Metropolis chain, we do not distinguish between them when referring to irreducibility.

Proof Let $\Omega = [q]^V$ and $P \in \mathbb{R}_{\geq 0}^{|\Omega| \times |\Omega|}$ denote the transition matrix for the *LocalMetropolis* chain. First, we show this chain is reversible and μ is stationary, by verifying the detailed balance equation:

$$\mu(X)P(X, Y) = \mu(Y)P(Y, X).$$

If two configurations X, Y are both infeasible, then $\mu(X) = \mu(Y) = 0$. If precisely one of X, Y is feasible, say X is feasible and Y is not, then $\mu(Y) = 0$ and X cannot move to Y since at least one edge cannot pass its check, which means $P(X, Y) = 0$. In both cases, the detailed balance equation holds.

Next, we suppose X, Y are both feasible. Consider a move in the *LocalMetropolis* chain. Let $\mathcal{C} \in \{0, 1\}^E$ be a Boolean vector that \mathcal{C}_e indicates whether edge $e \in E$ passes its check. We call $v \in V$ *non-restricted* by \mathcal{C} if $\mathcal{C}_e = 1$ for all e incident with v and v accepts the proposal; and call $v \in V$ *restricted* by \mathcal{C} if otherwise.

A move in the chain is completely determined by \mathcal{C} along with the proposed configurations $\sigma \in [q]^V$. Let $\Omega_{X \rightarrow Y}$ denote the set of pairs (σ, \mathcal{C}) with which X moves to Y , and $\Delta_{X, Y} = \{v \in V \mid X_v \neq Y_v\}$ the set of vertices on which X and Y disagree. Note that each $(\sigma, \mathcal{C}) \in \Omega_{X \rightarrow Y}$ satisfies:

- $\forall v \in \Delta_{X, Y}: \sigma_v = Y_v$ and v is non-restricted by \mathcal{C} ;
- $\forall v \notin \Delta_{X, Y}: \text{either } \sigma_v = X_v = Y_v \text{ or } v \text{ is restricted by } \mathcal{C}$.

Similar holds for $\Omega_{Y \rightarrow X}$, the set of (σ, \mathcal{C}) with which Y moves to X . Hence:

$$\frac{P(X, Y)}{P(Y, X)} = \frac{\sum_{(\sigma, \mathcal{C}) \in \Omega_{X \rightarrow Y}} \Pr(\sigma) \Pr(\mathcal{C} \mid \sigma, X)}{\sum_{(\sigma, \mathcal{C}) \in \Omega_{Y \rightarrow X}} \Pr(\sigma) \Pr(\mathcal{C} \mid \sigma, Y)}. \tag{7}$$

In order to verify the detailed balance equation, we construct a bijection $\phi_{X, Y} : \Omega_{X \rightarrow Y} \rightarrow \Omega_{Y \rightarrow X}$, and for every $(\sigma, \mathcal{C}) \in \Omega_{X \rightarrow Y}$, denoted $(\sigma', \mathcal{C}') = \phi_{X, Y}(\sigma, \mathcal{C})$, and show that

$$\frac{\Pr(\sigma) \Pr(\mathcal{C} \mid \sigma, X)}{\Pr(\sigma') \Pr(\mathcal{C}' \mid \sigma', Y)} = \frac{\mu(Y)}{\mu(X)}. \tag{8}$$

The detailed balance equation then follows from (7) and (8).

The bijection $(\sigma, \mathcal{C}) \xrightarrow{\phi_{X, Y}} (\sigma', \mathcal{C}')$ is constructed as follow:

- $\mathcal{C}' = \mathcal{C}$;
- for all v non-restricted by \mathcal{C} , since $(\sigma, \mathcal{C}) \in \Omega_{X \rightarrow Y}$ it must hold $\sigma_v = Y_v$, then set $\sigma'_v = X_v$;
- for all v restricted by \mathcal{C} , since $(\sigma, \mathcal{C}) \in \Omega_{X \rightarrow Y}$ it must hold $X_v = Y_v$, then set $\sigma'_v = \sigma_v$.

It can be verified that the $\phi_{X, Y}$ constructed in this way is indeed a bijection from $\Omega_{X \rightarrow Y}$ to $\Omega_{Y \rightarrow X}$. For any $(\sigma, \mathcal{C}) \in \Omega_{X \rightarrow Y}$ and the corresponding $(\sigma', \mathcal{C}') \in \Omega_{Y \rightarrow X}$, since $\mathcal{C}' =$

\mathcal{C} , in the following we will not specify whether v is (non-)restricted by \mathcal{C} or \mathcal{C}' but just say v is (non-)restricted, and the followings are satisfied:

- $\forall v \in \Delta_{X,Y}: \sigma_v = Y_v, \sigma'_v = X_v$ and v is non-restricted;
- $\forall v \notin \Delta_{X,Y}$: either $\sigma_v = \sigma'_v = X_v = Y_v$ or v is restricted and $\sigma_v = \sigma'_v$. In both cases, $\sigma_v = \sigma'_v$.

Then we have:

$$\frac{\Pr(\sigma)}{\Pr(\sigma')} = \frac{\prod_{v \in V} b_v(\sigma_v)}{\prod_{v \in V} b_v(\sigma'_v)} = \frac{\prod_{v: X_v \neq Y_v} b_v(\sigma_v)}{\prod_{v: X_v \neq Y_v} b_v(\sigma'_v)} = \frac{\prod_{v: X_v \neq Y_v} b_v(Y_v)}{\prod_{v: X_v \neq Y_v} b_v(X_v)} = \frac{\prod_{v \in V} b_v(Y_v)}{\prod_{v \in V} b_v(X_v)}. \tag{9}$$

Next, for each edge $e \in E$ we calculate the ratio $\frac{\Pr(\mathcal{C}_e | \sigma, X)}{\Pr(\mathcal{C}'_e | \sigma', Y)}$. There are two cases:

- If $\mathcal{C}_e = 0$ which means e does not pass its check, then

$$\Pr[\mathcal{C}_e = 0 | \sigma, X] = 1 - \tilde{A}_e(\sigma_u, \sigma_v) \tilde{A}_e(X_u, \sigma_v) \tilde{A}_e(\sigma_u, X_v)$$

and

$$\Pr[\mathcal{C}'_e = 0 | \sigma', Y] = 1 - \tilde{A}_e(\sigma'_u, \sigma'_v) \tilde{A}_e(Y_u, \sigma'_v) \tilde{A}_e(\sigma'_u, Y_v).$$

And both u and v are restricted by \mathcal{C} . By our construction of the bijection $\phi_{X,Y}$, we have $\sigma_u = \sigma'_u, \sigma_v = \sigma'_v, X_u = Y_u$, and $X_v = Y_v$. It follows that

$$\frac{\Pr[\mathcal{C}_e = 0 | \sigma, X]}{\Pr[\mathcal{C}'_e = 0 | \sigma', Y]} = \frac{A_e(Y_u, Y_v)}{A_e(X_u, X_v)} = 1.$$

- If $\mathcal{C}_e = 1$ which means e passes its check, then

$$\Pr[\mathcal{C}_e = 1 | \sigma, X] = \tilde{A}_e(\sigma_u, \sigma_v) \tilde{A}_e(X_u, \sigma_v) \tilde{A}_e(\sigma_u, X_v),$$

and

$$\Pr[\mathcal{C}'_e = 1 | \sigma', Y] = \tilde{A}_e(\sigma'_u, \sigma'_v) \tilde{A}_e(Y_u, \sigma'_v) \tilde{A}_e(\sigma'_u, Y_v).$$

There are three sub-cases according to whether vertices u and v are restricted:

1. Both u and v are restricted, in which case $\sigma_u = \sigma'_u, \sigma_v = \sigma'_v, X_u = Y_u, X_v = Y_v$.
2. Precisely one of $\{u, v\}$ is restricted, say v is restricted and u is non-restricted, in which case $\sigma_u = Y_u, \sigma'_u = X_u, \sigma_v = \sigma'_v$, and $X_v = Y_v$.
3. Both u and v are non-restricted, in which case $\sigma_u = Y_u, \sigma'_u = X_u, \sigma_v = Y_v, \sigma'_v = X_v$.

In all three sub-cases, the following identity can be verified:

$$\frac{\Pr[\mathcal{C}_e = 1 | \sigma, X]}{\Pr[\mathcal{C}'_e = 1 | \sigma', Y]} = \frac{\tilde{A}_e(Y_u, Y_v)}{\tilde{A}_e(X_u, X_v)} = \frac{A_e(Y_u, Y_v)}{A_e(X_u, X_v)}.$$

Since each edges passes its check independently, we have

$$\frac{\Pr(\mathcal{C} | \sigma, X)}{\Pr(\mathcal{C}' | \sigma', Y)} = \prod_{e=uv \in E} \frac{A_e(Y_u, Y_v)}{A_e(X_u, X_v)}. \tag{10}$$

Combining (9) and (10), for every $(\sigma, \mathcal{C}) \in \Omega_{X \rightarrow Y}$ and the corresponding $(\sigma', \mathcal{C}') \in \Omega_{Y \rightarrow X}$, we have:

$$\frac{\Pr[\sigma] \Pr[\mathcal{C} | \sigma, X]}{\Pr[\sigma'] \Pr[\mathcal{C}' | \sigma', Y]} = \prod_{v \in V} \frac{b_v(Y_v)}{b_v(X_v)} \prod_{e=uv \in E} \frac{A_e(Y_u, Y_v)}{A_e(X_u, X_v)} = \frac{\mu(Y)}{\mu(X)}.$$

This completes the verification of detailed balance equation and the proof of the reversibility of the chain with respect to stationary distribution μ .

Next, observe that the chain will never move from a feasible configuration to an infeasible one since at least one of the edge will not pass its check. By assumption (6), for all $X \in [q]^V$, no matter feasible or not, and for every $v \in V$ there must be a spin state $i \in [q]$ such that with positive probability v is successfully updated to spin state i . Note that once a vertex is successfully updated it satisfies and will keep satisfying all its local constraints. Therefore, the chain is absorbing to feasible configurations.

It is easy to observe that every feasible configuration is aperiodic, since it has self-loop transition, i.e. $P(X, X) > 0$ for all feasible X . In addition, any move $X \rightarrow Y$ between feasible configurations $X, Y \in \Omega$ in the single-site Markov chain with vertex v being updated, can be simulated by a move in the *LocalMetropolis* chain in which all the vertices u other than v propose their current spin state X_u and v proposes Y_v . Provided the irreducibility of the single-site Markov chain among all feasible configurations, the *LocalMetropolis* chain is also irreducible among all feasible configurations. Combining with the absorption towards feasible configurations and their aperiodicity, due to the Markov chain convergence theorem [43], $d_{TV}(\mu_{LM}, \mu)$ converges to 0 as $T \rightarrow \infty$. \square

4.2 The mixing of *LocalMetropolis* chain for graph colorings

Unlike the *LubyGlauber* chain, whose mixing rate is essentially due to the analysis of systematic scans. The mixing rate of *LocalMetropolis* chain is much more complicated to analyze. Here we analyze the mixing rate of the *LocalMetropolis* chain for proper q -colorings.

Given a graph $G(V, E)$, a q -coloring $\sigma \in [q]^V$ is proper if $\sigma_u \neq \sigma_v$ for all $uv \in E$. For this special MRF, the *LocalMetropolis* chain behaves simply as follows. Starting from an arbitrary coloring $X \in [q]^V$, not necessarily proper, in each step:

- **Propose:** each vertex v proposes a color $c_v \in [q]$ uniformly at random;
- **Local filter:** each vertex v rejects its proposal if there is a neighbor $u \in \Gamma(v)$ such that one of the followings occurs:
 1. (v proposed the neighbor’s current color) $c_v = X_u$;
 2. (v and the neighbor proposed the same color) $c_v = c_u$;
 3. (the neighbor proposed v ’s current color) $X_v = c_u$;
 otherwise, v accepts its proposal and updates its color X_v to c_v .

The first two filtering rules are sufficient to guarantee that the chain will never move to a “less proper” coloring. Although at first glance the third filtering rule looks redundant, it is necessary to guarantee the reversibility of the chain as well as the uniform stationary distribution.

It can be verified that when $q \geq \Delta + 2$, the condition (6) is satisfied and the single-site Glauber dynamics for proper q -coloring is irreducible, and hence the chain is mixing due to Theorem 5. The following theorem states a condition in the form $q \geq \alpha\Delta$ for the logarithmic mixing rate even for unbounded Δ and q . This proves Theorem 2.

Theorem 6 *If $q \geq \alpha\Delta$ for a constant $\alpha > 2 + \sqrt{2}$, the mixing rate of the LocalMetropolis chain for proper q -coloring on graphs with maximum degree at most $\Delta = \Delta(n) \geq 9$ is $\tau(\epsilon) = O(\log(\frac{n}{\epsilon}))$, where the constant factor in $O(\cdot)$ depends only on α but not on the maximum degree Δ .*

The theorem is proved by path coupling, a powerful engineering tool for coupling Markov chains. A coupling of a Markov chain on space Ω is a Markov chain $(X, Y) \rightarrow (X', Y')$ on space Ω^2 such that the transitions $X \rightarrow X'$ and $Y \rightarrow Y'$ individually follow the same transition rule as the original chain on Ω . For path coupling, we can construct a coupled Markov chain $(X, Y) \rightarrow (X', Y')$ for $X, Y \in [q]^V$ which differ at only one vertex. The chain mixes rapidly if the expected number of disagreeing vertices in (X', Y') is < 1 .

4.2.1 An ideal coupling

The $2 + \sqrt{2}$ threshold in Theorem 6 is due to an ideal coupling in the Δ -regular tree. Let \mathbb{T}_Δ denote the infinite Δ -regular tree rooted at v_0 . We assume that the current pair of colorings (X, Y) disagree only at the root v_0 and $X_u = Y_u \notin \{X_{v_0}, Y_{v_0}\}$ for all other vertices u in \mathbb{T}_Δ .

An ideal coupling can be constructed as follows in a breadth-first fashion: (1) the root v_0 proposes the same random color in both chains X, Y ; (2) each child u of the root proposes the same random color in both chains unless it proposed one of $\{X_{v_0}, Y_{v_0}\}$, in which case it switches the roles

of the two colors $\{X_{v_0}, Y_{v_0}\}$ in the Y chain; (3) for all other vertices u , it proposes the same random color in both chains unless its parent proposed different colors in the two chains, in which case u switches the roles of $\{X_{v_0}, Y_{v_0}\}$ in the Y chain. For this ideal coupling, by a calculation, it can be verified that for the root v_0 :

$$\Pr[X'_{v_0} \neq Y'_{v_0}] \leq 1 - \left(1 - \frac{\Delta}{q}\right) \left(1 - \frac{2}{q}\right)^\Delta$$

and for any non-root vertex u in \mathbb{T}_Δ at distance ℓ from v_0 :

$$\begin{aligned} \Pr[X'_u \neq Y'_u] &\leq \frac{1}{q} \left(1 - \frac{2}{q}\right)^{\Delta-1} \left(\frac{2}{q}\right)^{\ell-1} \\ &= \frac{1}{2} \left(1 - \frac{2}{q}\right)^{\Delta-1} \left(\frac{2}{q}\right)^\ell. \end{aligned}$$

The expected number of disagreeing vertices in (X', Y') is then bounded as

$$\begin{aligned} \Pr[X'_{v_0} \neq Y'_{v_0}] + \sum_{\substack{u \in \mathbb{T} \\ u \neq v_0}} \Pr[X'_u \neq Y'_u] \\ \leq 1 - \left(1 - \frac{\Delta}{q}\right) \left(1 - \frac{2}{q}\right)^\Delta + \frac{1}{2} \left(1 - \frac{2}{q}\right)^{\Delta-1} \sum_{\ell=1}^\infty \Delta^\ell \left(\frac{2}{q}\right)^\ell \\ = 1 - \left(1 - \frac{\Delta}{q}\right) \left(1 - \frac{2}{q}\right)^\Delta + \frac{\Delta}{q-2\Delta} \left(1 - \frac{2}{q}\right)^{\Delta-1}. \end{aligned}$$

The path coupling argument requires this quantity to be < 1 . For $q = \alpha^*\Delta$ and $\Delta \rightarrow \infty$, this quantity becomes $1 - e^{-2/\alpha^*} \left(1 - \frac{1}{\alpha^*} - \frac{1}{\alpha^*-2}\right)$, which is < 1 if $\alpha^* > 2 + \sqrt{2}$.

For general non-tree graphs $G(V, E)$ and arbitrary pairs of colorings (X, Y) which disagree at only one vertex, where X, Y may not even be proper, we essentially show that the above special pair of colorings (X, Y) on the infinite Δ -regular tree \mathbb{T}_Δ represent the worst case for path coupling. The analysis for this general case is quite involved. We first state the path coupling lemma with general metric.

Lemma 2 (Bubley and Dyer [6]) *Given a pre-metric, which is a connected undirected graph on configuration space Ω with positive edge weight such that every edge is a shortest path, let $\Phi(X, Y)$ be the length of the shortest path between two configurations $X, Y \in \Omega$. Suppose that there is a coupling $(X, Y) \rightarrow (X', Y')$ of the Markov chain defined only for the pair (X, Y) of configurations that are adjacent in the pre-metric, which satisfies that*

$$\mathbf{E}[\Phi(X', Y') \mid X, Y] \leq (1 - \delta)\Phi(X, Y),$$

for some $0 < \delta < 1$. Then the mixing rate of the Markov chain is bounded by

$$\tau(\epsilon) \leq \frac{\log(\text{diam}(\Omega)/\epsilon)}{\delta},$$

where $\text{diam}(\Omega)$ denotes the diameter of Ω in the pre-metric.

We use the following slightly modified pre-metric: A pair $(X, Y) \in \Omega = [q]^V$ is connected by an edge in the pre-metric if and only if X and Y differ at only one vertex, say v , and the edge-weight is given by $\text{deg}(v)$. This leads us to the following definition.

Definition 3 For any $X', Y' \in \Omega$, for $u \in V$, we define $\phi_u(X', Y') = \text{deg}(u)$ if $X'_u \neq Y'_u$ and $\phi_u(X', Y') = 0$ if otherwise; and for $S \subseteq V$, we define the distance between X' and Y' on S as

$$\Phi_S(X', Y') := \sum_{u \in S: X'_u \neq Y'_u} \phi_u(X', Y').$$

In addition, we denote $\Phi(X', Y') = \Phi_V(X', Y')$.

Clearly, the diameter of Ω in distance Φ has $\text{diam}(\Omega) \leq n\Delta$.

We prove the mixing rate in Theorem 6 for two separate regimes for q by using two different couplings. We define $\alpha^* \approx 3.634 \dots$ to be the positive root of $\alpha = 2e^{1/\alpha} + 1$.

Lemma 3 *If $q \geq \alpha\Delta + 3$ for a constant $\alpha > \alpha^*$, then $\tau(\epsilon) = O(\log(\frac{n}{\epsilon}))$.*

Lemma 4 *If $\alpha\Delta \leq q \leq 3.7\Delta + 3$ for $2 + \sqrt{2} < \alpha \leq 3.7$ and $\Delta \geq 9$, then $\tau(\epsilon) = O(\log(\frac{n}{\epsilon}))$.*

Theorem 6 follows by combining the two lemmas.

4.2.2 An easy local coupling for $q > 3.634\Delta + 3$

We first prove Lemma 3 by constructing a local coupling where the disagreement will not percolate outside its neighborhood. Let $X, Y \in [q]^V$ two q -colorings, not necessarily proper. Assume that X and Y disagree only at vertex $v_0 \in V$. The coupling $(X, Y) \rightarrow (X', Y')$ is constructed as follows:

- Each vertex $v \in V$ proposes the same random color in the two chains X and Y . Then (X', Y') is determined due to the transition rule of *LocalMetropolis* chain.

Next we show the path coupling condition:

$$\mathbf{E}[\Phi(X', Y') \mid X, Y] \leq (1 - \delta)\Phi(X, Y) = (1 - \delta) \text{deg}(v_0).$$

The following technical lemma is frequently applied in the analysis of this and next couplings.

Lemma 5 *If $q \geq a\Delta$, then for any integer $0 \leq d \leq \Delta$, $d \left(1 - \frac{a}{q}\right)^d \leq \Delta \left(1 - \frac{a}{q}\right)^\Delta$.*

Proof It is sufficient to show the function $d \left(1 - \frac{a}{q}\right)^d$ is monotone for integer $1 \leq d \leq \Delta$:

$$d \left(1 - \frac{a}{q}\right)^d - (d - 1) \left(1 - \frac{a}{q}\right)^{d-1} = \left(1 - \frac{a}{q}\right)^{d-1} \left(1 - \frac{ad}{q}\right),$$

which is nonnegative when $q \geq ad$. \square

Proof of Lemma 3. First, observe that if $v \notin \Gamma^+(v_0)$, where v_0 is the vertex at which X and Y disagree, then it always holds that $X'_v = Y'_v$, because all vertices in $\Gamma^+(v)$ are colored the same in X and Y and will propose the same random color in the two chains due to the coupling. Therefore, it is sufficient to consider the difference between X' and Y' in $\Gamma^+(v_0)$ and we have

$$\Phi(X', Y') = \Phi_{\Gamma^+(v_0)}(X', Y').$$

For each v , let $c_v \in [q]$ be the uniform random color proposed independently by v , which is identical in both chains by the coupling.

For the disagreeing vertex v_0 , it holds that $X'_{v_0} = Y'_{v_0}$ if v_0 accepts the proposal in both chains, which occurs when $c_{v_0} \notin \{X_u, Y_u : u \in \Gamma(v_0)\}$ and $\forall u \in \Gamma(v_0), c_u \notin \{X_{v_0}, Y_{v_0}, c_{v_0}\}$. Since X and Y disagree only at v_0 , we have

$$\Pr[X'_{v_0} = Y'_{v_0} \mid X, Y] \geq \left(1 - \frac{d_{v_0}}{q}\right) \left(1 - \frac{3}{q}\right)^{d_{v_0}}. \tag{11}$$

For each $u \in \Gamma(v_0)$, since $X_u = Y_u$, the event $X'_u \neq Y'_u$ occurs only when $c_u \in \{X_{v_0}, Y_{v_0}\}$ and $\forall w \in \Gamma(u), c_w \notin \{X_u, c_u\}$. Note that to guarantee $X'_u \neq Y'_u$ one must have $c_u \neq X_u$, thus

$$\forall u \in \Gamma(v_0) : \Pr[X'_u \neq Y'_u \mid X, Y] \leq \frac{2}{q} \left(1 - \frac{2}{q}\right)^{d_u}. \tag{12}$$

Combining (11) and (12) together and due to linearity of expectation, we have

$$\begin{aligned} \mathbf{E}[\Phi(X', Y') \mid X, Y] &= \sum_{u \in V} \mathbf{E}[\phi_u(X', Y') \mid X, Y] \\ &= \sum_{u \in \Gamma^+(v_0)} d_u \Pr[X'_u \neq Y'_u \mid X, Y] \\ &\leq d_{v_0} \left[1 - \left(1 - \frac{d_{v_0}}{q}\right) \left(1 - \frac{3}{q}\right)^{d_{v_0}} \right] \\ &\quad + \frac{2}{q} \sum_{u \in \Gamma(v_0)} d_u \left(1 - \frac{2}{q}\right)^{d_u} \\ &\leq d_{v_0} \left[1 - \left(1 - \frac{\Delta}{q}\right) \left(1 - \frac{3}{q}\right)^\Delta + \frac{2\Delta}{q} \left(1 - \frac{2}{q}\right)^\Delta \right], \end{aligned}$$

where the last inequality is due to the monotonicity stated in Lemma 5.

The path coupling condition is satisfied when

$$\left(1 - \frac{\Delta}{q}\right) \left(1 - \frac{3}{q}\right)^\Delta - \frac{2\Delta}{q} \left(1 - \frac{2}{q}\right)^\Delta \geq \delta. \tag{13}$$

For $q = \alpha^* \Delta$ and $\Delta \rightarrow \infty$, then the LHS becomes $\left(1 - \frac{1}{\alpha^*}\right) e^{-3/\alpha^*} - \frac{2}{\alpha^*} e^{-2/\alpha^*}$, which is 0 when α^* is the positive root of $\alpha^* = 2e^{1/\alpha^*} + 1$.

Furthermore, for $\Delta \geq 1$ and $q \geq \alpha \Delta + 3$, the LHS become:

$$\begin{aligned} &\left(1 - \frac{3}{q}\right)^\Delta \left[1 - \frac{\Delta}{q} - \frac{2\Delta}{q} \left(1 + \frac{1}{q-3}\right)^\Delta\right] \\ &\geq \left(1 - \frac{3}{\alpha \Delta + 3}\right)^\Delta \left[1 - \frac{1}{\alpha} - \frac{2}{\alpha} \left(1 + \frac{1}{\alpha \Delta}\right)^\Delta\right] \\ &\geq \frac{e^{-3/\alpha}}{\alpha} (\alpha - 2e^{1/\alpha} - 1), \end{aligned}$$

which is a positive constant independent of Δ when $\alpha > \alpha^*$.

Therefore, when $\alpha > \alpha^*$, there is a constant $\delta > 0$ which depends only on α , such that for all $\Delta \geq 1$ and $q \geq \alpha \Delta + 3$, the inequality (13) is satisfied, which by Lemma 2, gives us $\tau(\epsilon) = O\left(\log\left(\frac{n}{\epsilon}\right)\right)$.

4.2.3 A global coupling for $(2 + \sqrt{2})\Delta < q \leq 3.7\Delta + 3$

Next, we prove Lemma 4 and bound the mixing rate when $(2 + \sqrt{2})\Delta < q \leq 3.7\Delta + 3$. This is done by a global coupling where the disagreement may percolate to the entire graph, whose construction and analysis is substantially more sophisticated than the previous local coupling. Although this sophistication only improves the threshold for q in Lemma 3 by a small constant factor, the effort is worthwhile because it helps us to approach the threshold of the ideal coupling discussed in Sect. 4.2.1 and shows that the infinite Δ -regular tree \mathbb{T}_Δ represents the worst case for path coupling. And curiously, the extremity of this worst case only holds when q is also properly upper bounded, say $q \leq 3.7\Delta + 3$, whereas the mixing rate for larger q was guaranteed by Lemma 3.

Let $v_0 \in V$ be a vertex and $X, Y \in [q]^V$ any two q -colorings (not necessarily proper) which disagree only at v_0 . The coupling $(X, Y) \rightarrow (X', Y')$ of the *LocalMetropolis* chain is constructed by coupling (c^X, c^Y) , where $c^X, c^Y \in [q]^V$ are the respective vector of proposed colors in the two chains X and Y . For each $v \in V$, the (c_v^X, c_v^Y) is sampled from one of the two following joint distributions:

- **consistent:** $c_v^X = c_v^Y$ and is uniformly distributed over $[q]$;

- **permuted:** c_v^X is uniform in $[q]$ and $c_v^Y = \phi(c_v^X)$ where $\phi : [q] \rightarrow [q]$ is a bijection defined as that $\phi(X_{v_0}) = Y_{v_0}$, $\phi(Y_{v_0}) = X_{v_0}$, and $\phi(x) = x$ for all $x \notin \{X_{v_0}, Y_{v_0}\}$.

Note that for all $u \neq v_0$ we have $X_u = Y_u$, and if further $X_u \in \{X_{v_0}, Y_{v_0}\}$, we say the vertices $w \in \Gamma^+(u) \setminus \{v_0\}$ are *blocked* by u , and all other $u \neq v_0$ is *unblocked*. The special vertex v_0 is neither blocked nor unblocked. We denote by $\Gamma^B(v)$ and $\Gamma^U(v)$ the respective sets of blocked and unblocked neighbors of vertex v and let $b_v = |\Gamma^B(v)|$.

The coupling (c^X, c^Y) of proposed colors is constructed by the following recursive procedure:

- Initially, for the disagreeing vertex v_0 , $(c_{v_0}^X, c_{v_0}^Y)$ is sampled consistently in the two chains.
- For each unblocked $u \in \Gamma(v_0)$, the (c_u^X, c_u^Y) is sampled independently (of other vertices) from the permuted distribution.
- Let $\mathcal{S} \subseteq V$ denote the current set of vertices v such that (c_v^X, c_v^Y) has been sampled, and $\mathcal{S}^\neq \subseteq \mathcal{S}$ the set of vertices v with (c_v^X, c_v^Y) sampled inconsistently as $c_v^X \neq c_v^Y$. We abuse the notation and use $\partial \mathcal{S}^\neq = \{\text{unblocked } u \notin \mathcal{S} \mid \exists uv \in E, \text{ s.t. } v \in \mathcal{S}^\neq\}$ to denote the unblocked un-sampled vertex boundary of \mathcal{S}^\neq . If such $\partial \mathcal{S}^\neq$ is non-empty, then all $u \in \partial \mathcal{S}^\neq$ sample the respective (c_u^X, c_u^Y) independently from the permuted distribution and join the \mathcal{S} simultaneously. Grow \mathcal{S}^\neq according to the results of sampling. Repeat this step until the current $\partial \mathcal{S}^\neq$ is empty and thus \mathcal{S} is stabilized.
- For all remaining vertices v , (c_v^X, c_v^Y) is sampled independently and consistently.

This procedure is in fact a Galton-Watson branching process starting from root v_0 . The blocked-ness of each vertex is determined by the current X and Y . The \mathcal{S} grows from the root by a percolation of disagreement $c_v^X \neq c_v^Y$ added in a breadth-first order.

It is easy to see that each individual c_v^X or c_v^Y is uniformly distributed over $[q]$ and is independent of c_u^X or c_u^Y for all other $u \neq v$ (although the joint distributions (c_v^X, c_v^Y) may be dependent of each other). Therefore, the (c^X, c^Y) is a valid coupling of proposed colors.

A walk $\mathcal{P} = (v_0, v_1, \dots, v_\ell)$ in $G(V, E)$ is called a *strongly self-avoiding walk (SSAW)* if \mathcal{P} is a simple path in G and $v_i v_j$ is not an edge in G for any $0 < i + 1 < j \leq \ell$. An SSAW $\mathcal{P} = (v_0, v_1, \dots, v_\ell)$ is said to be a *path of disagreement* with respect to (c^X, c^Y) if $(c_{v_i}^X, c_{v_i}^Y)$, $v_i \in \mathcal{P}$ are sampled in the order along the path \mathcal{P} from $i = 0$ to ℓ , and $c_{v_i}^X \neq c_{v_i}^Y$ for all $1 \leq i \leq \ell$. For any specific SSAW $\mathcal{P} = (v_0, v_1, \dots, v_\ell)$ through unblocked vertices v_1, v_2, \dots, v_ℓ , by the chain rule

$\Pr[\mathcal{P}$ is a path of disagreement]

$$\leq \prod_{i=1}^{\ell} \Pr \left[c_{v_i}^X \in \{X_{v_0}, Y_{v_0}\} \right] = \left(\frac{2}{q} \right)^\ell. \tag{14}$$

Proposition 2 For any vertex $u \neq v_0$, the event $c_u^X \neq c_u^Y$ occurs only if there is a strongly self-avoiding walk (SSAW) $\mathcal{P} = (v_0, v_1, \dots, v_\ell)$ from v_0 to $v_\ell = u$ through unblocked vertices v_1, v_2, \dots, v_ℓ such that \mathcal{P} is a path of disagreement.

Proof By the coupling, $c_u^X \neq c_u^Y$ only when (c_u^X, c_u^Y) is sampled from the permuted distribution and it must hold that $\{c_u^X, c_u^Y\} = \{X_{v_0}, Y_{v_0}\}$. This means that u itself must be unblocked.

At the time when (c_u^X, c_u^Y) is being sampled, there must exist a neighbor $w \in \Gamma(u)$ such that either (1) $w = v_0$ or (2) $w \in \mathcal{S}^\neq$, which means that $c_w^X \neq c_w^Y$, $\{c_w^X, c_w^Y\} = \{X_{v_0}, Y_{v_0}\}$ was sampled before (c_u^X, c_u^Y) , and vertex w is unblocked. If it is the latter case, we repeat this argument for w recursively until v_0 is reached. This will give us a path $\mathcal{P} = (v_0, v_1, \dots, v_\ell)$ from v_0 to $u = v_\ell$ through unblocked vertices v_1, \dots, v_ℓ such that for all $1 \leq i \leq \ell$, $(c_{v_i}^X, c_{v_i}^Y)$ are sampled in that order, $c_{v_i}^X \neq c_{v_i}^Y$ and $\{c_{v_i}^X, c_{v_i}^Y\} = \{X_{v_0}, Y_{v_0}\}$. Thus, \mathcal{P} is a path of disagreement through unblocked vertices. Note that this path $\mathcal{P} = (v_0, v_1, \dots, v_\ell)$ must be a strongly self-avoiding. To the contrary assume that \mathcal{P} is not strongly self-avoiding and there exist $0 \leq i, j \leq \ell$ such that $i < j - 1$ and $v_i v_j$ is an edge. In this case, right after $c_{v_i}^X \neq c_{v_i}^Y$ being sampled and v_i joining \mathcal{S}^\neq , v_{i+1} and v_j must be both in $\partial\mathcal{S}^\neq$ because they are both unblocked un-sampled neighbors of v_i then. And due to our construction of coupling, the $(c_{v_{i+1}}^X, c_{v_{i+1}}^Y)$ and $(c_{v_j}^X, c_{v_j}^Y)$ are sampled and v_{i+1}, v_j join \mathcal{S} simultaneously, which contradict that $(c_{v_j}^X, c_{v_j}^Y)$ is sampled after $(c_{v_{i+1}}^X, c_{v_{i+1}}^Y)$ along the path. Therefore, \mathcal{P} is an SSAW through unblocked vertices and is also a path of disagreement. \square

The coupled next step (X', Y') is determined by the current (X, Y) and the coupled proposed colors (c^X, c^Y) .

Proposition 3 For any vertex $u \neq v_0$, the event $X'_u \neq Y'_u$ occurs only if $c_u^X, c_u^Y \in \{X_{v_0}, Y_{v_0}\}$. Furthermore, for any unblocked vertex $u \neq v_0$, the event $X'_u \neq Y'_u$ occurs only if $c_u^X \neq c_u^Y$.

Proof We pick any $u \neq v_0$. Assume by contradiction that $c_u^X = c_u^Y \notin \{X_{v_0}, Y_{v_0}\}$. Note that this covers all possible contradicting cases to that $c_u^X, c_u^Y \in \{X_{v_0}, Y_{v_0}\}$, because $c_u^X \neq c_u^Y$ occurs only when $c_u^X, c_u^Y \in \{X_{v_0}, Y_{v_0}\}$.

We then show for every edge uw incident to u , the followings hold:

$$c_u^X = c_w^X \text{ if and only if } c_u^Y = c_w^Y, \tag{15}$$

$$X_u = c_w^X \text{ if and only if } Y_u = c_w^Y, \tag{16}$$

$$c_u^X = X_w \text{ if and only if } c_u^Y = Y_w. \tag{17}$$

With (15), (16) and (17), each edge uw passes the check in chain X if and only if it passes the check in chain Y . Combining with the fact that $X_u = Y_u$ for all $u \neq v_0$, this implies $X'_u = Y'_u$, a contradiction.

We then verify (15), (16) and (17):

- If $X_u = Y_u \in \{X_{v_0}, Y_{v_0}\}$, then for every neighbor $w \in \Gamma(u)$, either w is blocked or $w = v_0$. In both cases $c_w^X = c_w^Y$ is sampled consistently, this implies (15) and (16), because $c_u^X = c_u^Y$ and $X_u = Y_u$. And it holds that either $\{X_w, Y_w\} = \{X_{v_0}, Y_{v_0}\}$ (in case of $w = v_0$) or $X_w = Y_w$ (in case of $w \neq v_0$), this implies (17) because $c_u^X = c_u^Y \notin \{X_{v_0}, Y_{v_0}\}$.
- If $X_u = Y_u \notin \{X_{v_0}, Y_{v_0}\}$. For each neighbor $w \in \Gamma(u)$, it holds that either $\{c_w^X, c_w^Y\} = \{X_{v_0}, Y_{v_0}\}$ or $c_w^X = c_w^Y$, because the event $c_w^X \neq c_w^Y$ happens if and only if $\{c_w^X, c_w^Y\} = \{X_{v_0}, Y_{v_0}\}$ due to the coupling. Recall that $c_u^X = c_u^Y \notin \{X_{v_0}, Y_{v_0}\}$ and $X_u = Y_u \notin \{X_{v_0}, Y_{v_0}\}$, this implies (15) and (16). And it holds that either $\{X_w, Y_w\} = \{X_{v_0}, Y_{v_0}\}$ (in case of $w = v_0$) or $X_w = Y_w$ (in case of $w \neq v_0$), this implies (17) because $c_u^X = c_u^Y \notin \{X_{v_0}, Y_{v_0}\}$.

For an unblocked vertex $u \neq v_0$, assume $X'_u \neq Y'_u$. By above argument, we must have $c_u^X, c_u^Y \in \{X_{v_0}, Y_{v_0}\}$. We then show that $c_u^X \neq c_u^Y$. By contradiction, we assume $c_u^X = c_u^Y$, since $c_u^X, c_u^Y \in \{X_{v_0}, Y_{v_0}\}$, the (c_u^X, c_u^Y) must be sampled from the consistent distribution. And since u is unblocked and $u \neq v_0$, the (c_u^X, c_u^Y) is sampled from the consistent distribution only when for all neighbors $w \in \Gamma(u)$, $w \neq v_0$ (which means $X_w = Y_w$) and $c_w^X = c_w^Y$. In summary, $X_u = Y_u$, $c_u^X = c_u^Y$, and $X_w = Y_w$, $c_w^X = c_w^Y$ for all neighbors $w \in \Gamma(u)$, which guarantees that $X'_u = Y'_u$, a contradiction. Therefore, we also show that for any unblocked $u \neq v_0$, $X'_u \neq Y'_u$ only if $c_u^X \neq c_u^Y$. \square

We then analyze the probability of $X'_u \neq Y'_u$ for each vertex $u \in V$.

Lemma 6 For the vertex v_0 at which the q -colorings $X, Y \in [q]^V$ disagree,

$$\Pr[X'_{v_0} = Y'_{v_0} \mid X, Y] \geq \left(1 - \frac{\Delta}{q}\right) \left(1 - \frac{2}{q}\right)^\Delta \left(1 - \frac{1}{q-2}\right)^{b_{v_0}}.$$

Proof The event $X'_{v_0} = Y'_{v_0}$ occurs if v_0 accepts the proposal, which happens if the following events occur simultaneously:

- $c_{v_0}^X \notin \{X_u \mid u \in \Gamma(v_0)\}$ (and hence $c_{v_0}^Y \notin \{Y_u \mid u \in \Gamma(v_0)\}$ by the coupling $c_{v_0}^Y = c_{v_0}^X$ and the fact that $X_u = Y_u$ for $u \neq v_0$). This occurs with probability at least $\frac{q-d_{v_0}}{q}$.

- For all unblocked neighbors $u \in \Gamma^U(v_0)$, it must have $c_u^X \notin \{X_{v_0}, c_{v_0}^X\}$ and $c_u^Y \notin \{Y_{v_0}, c_{v_0}^Y\}$. This occurs with probability at least $\left(1 - \frac{2}{q}\right)^{d_{v_0} - b_{v_0}}$ conditioning on any choice of $c_{v_0}^X = c_{v_0}^Y$.
- For all blocked neighbors $w \in \Gamma^B(v_0)$, it must have $c_w^X \notin \{c_{v_0}^X, X_{v_0}, Y_{v_0}\}$ (and hence $c_w^Y \notin \{c_{v_0}^Y, X_{v_0}, Y_{v_0}\}$ due to the coupling $c_w^Y = c_w^X$). This occurs with probability at least $\left(1 - \frac{3}{q}\right)^{b_{v_0}}$ conditioning on any choice of $c_{v_0}^X = c_{v_0}^Y$ and independent of unblocked neighbors $u \in \Gamma^U(v_0)$.

Thus the following is obtained by the chain rule:

$$\begin{aligned} & \Pr[X'_{v_0} = Y'_{v_0} \mid X, Y] \\ & \geq \frac{q - d_{v_0}}{q} \left(1 - \frac{2}{q}\right)^{d_{v_0} - b_{v_0}} \left(1 - \frac{3}{q}\right)^{b_{v_0}} \\ & \geq \left(1 - \frac{\Delta}{q}\right) \left(1 - \frac{2}{q}\right)^\Delta \left(1 - \frac{1}{q-2}\right)^{b_{v_0}}, \end{aligned}$$

where the last inequality is due to the monotonicity stated in Lemma 5. \square

Lemma 7 For any unblocked vertex $u \neq v_0$, it holds that

$$\begin{aligned} & \Pr[X'_u \neq Y'_u \mid X, Y] \\ & \leq \frac{1}{q} \left(1 - \frac{2}{q}\right)^{d_u - 1} \left[2 - \left(1 - \frac{1}{q-2}\right)^{b_u}\right] \\ & \times \sum_{\substack{\text{unblocked SSAW} \\ \mathcal{P} \text{ from } v_0 \text{ to } u}} \left(\frac{2}{q}\right)^{\ell(\mathcal{P}) - 1}, \end{aligned} \tag{18}$$

where the sum enumerates all strongly self-avoiding walks (SSAW) $\mathcal{P} = (v_0, v_1, \dots, v_\ell)$ from v_0 to $v_\ell = u$ over unblocked vertices $v_1, v_2, \dots, v_\ell = u$, and $\ell(\mathcal{P}) = \ell$ denotes the length of the walk \mathcal{P} .

Proof Due to Proposition 3, for unblocked $u \neq v_0$, the event $X'_u \neq Y'_u$ occurs only if $c_u^X \neq c_u^Y$ and u accepts its proposal in at least one chain among X, Y . Observe that any edge uv between unblocked vertices u, v either passes the check in both chains X, Y or does not pass the check in both chains. Therefore, the event $X'_u \neq Y'_u$ occurs for an unblocked $u \neq v_0$ only if the following events occurs simultaneously:

- $c_u^X \neq c_u^Y$, which according to Proposition 2, occurs only if there is a SSAW $\mathcal{P} = (v_0, v_1, \dots, v_\ell)$ from v_0 to $v_\ell = u$ through unblocked vertices v_1, \dots, v_ℓ such that \mathcal{P} is a path of disagreement;
- for all unblocked neighbors $w \in \Gamma^U(u)$, the edge uw passes the check, which means $c_w^X \notin \{c_u^X, X_u\}$ (and meanwhile $c_w^Y \notin \{c_u^Y, Y_u\}$ by coupling) for all $w \in \Gamma^U(u)$;

- all blocked neighbors $w \in \Gamma^B(u)$ passes the check in at least one chains among X, Y , which means either $c_w^X \notin \{c_u^X, X_u\}$ for all $w \in \Gamma^B(u)$ or $c_w^Y \notin \{c_u^Y, Y_u\}$ for all $w \in \Gamma^B(u)$.

More specifically, these events occur only if:

- there is a SSAW $\mathcal{P} = (v_0, v_1, \dots, v_\ell)$ from v_0 to $v_\ell = u$ through unblocked vertices v_1, \dots, v_ℓ such that $c_{v_i}^X \in \{X_{v_0}, Y_{v_0}\}$ for $1 \leq i \leq \ell - 1$, which occurs with probability $\left(\frac{2}{q}\right)^{\ell - 1}$;
- if $u \in \Gamma(v_0)$, then $c_u^X = Y_{v_0}$ (and meanwhile $c_u^Y = X_{v_0}$ by coupling), and if $u \notin \Gamma(v_0)$, $c_u^X \in \{X_{v_0}, Y_{v_0}\} \setminus \{c_{v_{\ell-1}}^X\}$ (and meanwhile $c_u^Y \in \{X_{v_0}, Y_{v_0}\} \setminus \{c_{v_{\ell-1}}^Y\}$ by coupling), which in either case, occurs with probability $\frac{1}{q}$ conditioning on $(c_{v_{\ell-1}}^X, c_{v_{\ell-1}}^Y)$;
- $c_w^X \notin \{c_u^X, X_u\}$ (and meanwhile $c_w^Y \notin \{c_u^Y, Y_u\}$ by coupling) for all unblocked $w \in \Gamma^U(u) \setminus \{v_{\ell-1}\}$, which occurs with probability $\left(1 - \frac{2}{q}\right)^{d_u - b_u - 1}$ conditioning on c_u^X ;
- either $c_w^X \notin \{c_u^X, X_u\}$ for all $w \in \Gamma^B(u)$ or $c_w^Y \notin \{c_u^Y, Y_u\}$ for all $w \in \Gamma^B(u)$, which occurs with probability at most $\left[2 \left(1 - \frac{2}{q}\right)^{b_u} - \left(1 - \frac{3}{q}\right)^{b_u}\right]$ conditioning on (c_u^X, c_u^Y) by the principle of inclusion-exclusion.

Take the union bound over all SSAW $\mathcal{P} = (v_0, v_1, \dots, v_\ell)$ through unblocked vertices $v_1, \dots, v_\ell = u$. Due to the strongly-avoiding property, it is safe to apply the chain rule for every \mathcal{P} . We have:

$$\begin{aligned} & \Pr[X'_u \neq Y'_u \mid X, Y] \\ & \leq \sum_{\substack{\text{unblocked SSAW} \\ \mathcal{P} \text{ from } v_0 \text{ to } u}} \left(\left(\frac{2}{q}\right)^{\ell(\mathcal{P}) - 1} \left(\frac{1}{q}\right) \left(1 - \frac{2}{q}\right)^{d_u - b_u - 1}\right. \\ & \quad \times \left. \left[2 \left(1 - \frac{2}{q}\right)^{b_u} - \left(1 - \frac{3}{q}\right)^{b_u}\right]\right) \\ & = \frac{1}{q} \left(1 - \frac{2}{q}\right)^{d_u - 1} \left[2 - \left(1 - \frac{1}{q-2}\right)^{b_u}\right] \sum_{\substack{\text{unblocked SSAW} \\ \mathcal{P} \text{ from } v_0 \text{ to } u}} \left(\frac{2}{q}\right)^{\ell(\mathcal{P}) - 1}. \end{aligned}$$

\square

Lemma 8 For any blocked vertex $u \neq v_0$, it holds that

$$\begin{aligned} & \Pr[X'_u \neq Y'_u \mid X, Y] \\ & \leq \frac{1}{q} \left(1 - \frac{2}{q}\right)^{d_u - 1} \sum_{\substack{\text{SSAW } \mathcal{P} \text{ from } v_0 \text{ to } u \\ \text{with only } u \text{ blocked}}} \left(\frac{2}{q}\right)^{\ell(\mathcal{P}) - 1}, \end{aligned} \tag{19}$$

where the sum enumerates all the strongly self-avoiding walks (SSAW) $\mathcal{P} = (v_0, v_1, \dots, v_\ell)$ from v_0 to $v_\ell = u$

through unblocked vertices $v_1, \dots, v_{\ell-1}$, and $\ell(\mathcal{P}) = \ell$ denotes the length of the walk \mathcal{P} .

Proof By the coupling, any blocked vertex $u \in V$ proposes consistently in the two chains, thus $c_u^X = c_u^Y$. And we have $X_u = Y_u$ for $u \neq v_0$.

We first consider v_0 's blocked neighbors $u \in \Gamma^B(v_0)$. There are two cases for such vertex u :

- $X_u = Y_u \notin \{X_{v_0}, Y_{v_0}\}$. Since vertex u is blocked, there must exist a vertex $w_0 \in \Gamma(u) \setminus \{v_0\}$, such that $X_{w_0} = Y_{w_0} \in \{X_{v_0}, Y_{v_0}\}$. Without loss of generality, suppose $X_{w_0} = Y_{w_0} = X_{v_0}$ (and the case $X_{w_0} = Y_{w_0} = Y_{v_0}$ follows by symmetry). By Proposition 3, $X'_u \neq Y'_u$ only if $c_u^X = c_u^Y \in \{X_{v_0}, Y_{v_0}\}$. Note that if $c_u^X = c_u^Y = X_{v_0}$, then the edge uw_0 cannot pass the check in both chains, hence $X'_u = Y'_u$, a contradiction. So we must have $c_u^X = c_u^Y = Y_{v_0}$, in which case edge v_0u cannot pass the check in chain Y , thus the event $X'_u \neq Y'_u$ occurs only when u accepts the proposal in chain X , which happens only if for all $w \in \Gamma(u)$, $c_w^X \notin \{c_u^X, X_u\}$. Remember that we already have $c_u^X = Y_{v_0} \neq X_u$ and note that all vertices in chain X propose independently, therefore $X'_u \neq Y'_u$ occurs with probability at most $\frac{1}{q} \left(1 - \frac{2}{q}\right)^{d_u}$.
- $X_u = Y_u \in \{X_{v_0}, Y_{v_0}\}$. Without loss of generality, suppose $X_u = Y_u = X_{v_0}$ (and the case $X_u = Y_u = Y_{v_0}$ follows by symmetry). By Proposition 3, $X'_u \neq Y'_u$ only if $c_u^X = c_u^Y \in \{X_{v_0}, Y_{v_0}\}$. If $c_u^X = c_u^Y = X_{v_0}$, the proposal and the current color of u are the same in two chains, hence $X'_u = Y'_u$, a contradiction. So we must have $c_u^X = c_u^Y = Y_{v_0}$, in which case the edge uv_0 cannot pass the check in chain Y , thus event $X'_u \neq Y'_u$ occurs only if vertex u accepts the proposal in chain X , which happens only if for all $w \in \Gamma(u)$, $c_w^X \notin \{c_u^X, X_u\} = \{X_{v_0}, Y_{v_0}\}$. Remember that we already have $c_u^X = Y_{v_0}$ and note that all vertices in chain X propose independently, therefore $X'_u \neq Y'_u$ occurs with probability at most $\frac{1}{q} \left(1 - \frac{2}{q}\right)^{d_u}$.

Hence, for all $u \in \Gamma^B(v_0)$, we have:

$$\Pr[X'_u \neq Y'_u \mid X, Y] \leq \frac{1}{q} \left(1 - \frac{2}{q}\right)^{d_u} \leq \frac{1}{q} \left(1 - \frac{2}{q}\right)^{d_u-1}.$$

The walk $\mathcal{P} = (v_0, u)$ is a strongly self-avoiding walk (SSAW) from v_0 to u with only u blocked. Therefore (19) is proved for blocked vertices $u \in \Gamma^B(v_0)$.

Now we consider the general blocked vertices $u \notin \Gamma^+(v_0)$. Assume that $X'_u \neq Y'_u$.

If u is blocked by itself, i.e. $X_u = Y_u \in \{X_{v_0}, Y_{v_0}\}$, then all the vertices $w \in \Gamma^+(u)$ are blocked and hence propose consistently, and for $u \notin \Gamma^+(v_0)$ all neighbors w have $X_w =$

Y_w , so we must have $X'_u = Y'_u$. Thus $\Pr[X'_u \neq Y'_u \mid X, Y] = 0$ and (19) holds trivially.

If otherwise u is not blocked by itself, i.e. $X_u = Y_u \notin \{X_{v_0}, Y_{v_0}\}$, then u must be blocked by one of its neighbors $w_0 \in \Gamma(u)$ such that $X_{w_0} = Y_{w_0} \in \{X_{v_0}, Y_{v_0}\}$. By Proposition 3, $X'_u \neq Y'_u$ only if $c_u^X = c_u^Y \in \{X_{v_0}, Y_{v_0}\}$. We must have $c_u^X \neq X_{w_0}$, because if otherwise $c_u^X = X_{w_0}$, together with that $c_u^Y = Y_{w_0}$ which is due to that $c_u^X = c_u^Y$ and $X_{w_0} = Y_{w_0}$, the edge uw_0 cannot pass the check in both chains, giving us $X'_u = Y'_u$, a contradiction.

For the following, we assume $c_u^X = c_u^Y \in \{X_{v_0}, Y_{v_0}\}$ and $c_u^X \neq X_{w_0}$, therefore $c_u^Y \neq Y_{w_0}$ because $c_u^X = c_u^Y$ and $X_{w_0} = Y_{w_0}$. We claim that u must have an unblocked neighbor $w^* \in \Gamma(u)$ such that $c_{w^*}^X \neq c_{w^*}^Y$ because if otherwise for all the vertices $w \in \Gamma^+(u)$, the consistencies $c_w^X = c_w^Y$ and $X_w = Y_w$ hold, giving us $X'_u = Y'_u$, a contradiction. Therefore, there is a neighbor $w^* \in \Gamma(u)$ such that $c_{w^*}^X \neq c_{w^*}^Y$, which by Proposition 2, means that there is a strongly self-avoiding walk (SSAW) $\mathcal{P} = (v_0, v_1, \dots, v_\ell)$ from v_0 to $v_\ell = u$ through unblocked $v_1, v_2, \dots, v_{\ell-1} = w^*$ such that $\mathcal{P}' = (v_0, v_1, \dots, v_{\ell-1})$ is a path of disagreement. Fix any SSAW $\mathcal{P} = (v_0, v_1, \dots, v_\ell)$ from v_0 to $v_\ell = u$ with only u blocked. By Proposition 2:

- $\mathcal{P}' = (v_0, v_1, \dots, v_{\ell-1})$ is a path of disagreement with probability at most $\left(\frac{2}{q}\right)^{\ell-1}$.

As argued above, assuming $X'_u \neq Y'_u$ we must have

- $c_u^X \in \{X_{v_0}, Y_{v_0}\} \setminus \{X_{w_0}\}$ (and $c_u^Y = c_u^X$ due to the coupling), which occurs with probability $\frac{1}{q}$ conditioning on that \mathcal{P}' is a path of disagreement.

As argued above, we have $\{c_{v_{\ell-1}}^X, c_{v_{\ell-1}}^Y\} = \{c_u^X, X_{w_0}\} = \{c_u^Y, Y_{w_0}\} = \{X_{v_0}, Y_{v_0}\}$. Without loss of generality, suppose $c_{v_{\ell-1}}^X = c_u^X = c_u^Y$ and $c_{v_{\ell-1}}^Y = X_{w_0} = Y_{w_0}$ (and the case $c_{v_{\ell-1}}^X = X_{w_0} = Y_{w_0}$ and $c_{v_{\ell-1}}^Y = c_u^X = c_u^Y$ follows by symmetry). Then edge $uv_{\ell-1}$ cannot pass the check in chain X because $c_{v_{\ell-1}}^X = c_u^X$. Then the event $X'_u \neq Y'_u$ occurs only if vertex u accepts the proposal in chain Y , which happens only if

- $c_w^Y \notin \{Y_u, c_u^Y\}$ for all $w \in \Gamma(u) \setminus \{v_{\ell-1}\}$. Recall that $Y_u \neq c_u^Y$. Since \mathcal{P} is a strongly self-avoiding, we have $w \notin \mathcal{P}$ for all $w \in \Gamma(u) \setminus \{v_{\ell-1}\}$. And the proposals are mutually independent in one chain. Condition on previous events, this probability is at most $\left(1 - \frac{2}{q}\right)^{d_u-1}$.

By the union bound over all SSAW \mathcal{P} from v_0 to u with u being the only blocked vertex, and the chain rule for every \mathcal{P} , we have

$$\begin{aligned} & \Pr[X'_u \neq Y'_u \mid X, Y] \\ & \leq \frac{1}{q} \left(1 - \frac{2}{q}\right)^{d_u-1} \sum_{\substack{\text{SSAW } \mathcal{P} \text{ from } v_0 \text{ to } u \\ \text{with only } u \text{ blocked}}} \left(\frac{2}{q}\right)^{\ell(\mathcal{P})-1}. \end{aligned}$$

This proves (19). □

We then verify the path coupling condition: for some constant $\delta > 0$,

$$\mathbf{E}[\Phi(X', Y') \mid X, Y] \leq (1 - \delta)\Phi(X, Y). \tag{20}$$

By the linearity of expectation,

$$\begin{aligned} & \mathbf{E}[\Phi(X', Y') \mid X, Y] \\ & = \sum_{u \in V} \mathbf{E}[\phi_u(X', Y') \mid X, Y] \\ & = d_{v_0} \Pr[X'_{v_0} \neq Y'_{v_0} \mid X, Y] \\ & \quad + \sum_{\substack{\text{unblocked} \\ u \neq v_0}} d_u \Pr[X'_u \neq Y'_u \mid X, Y] \\ & \quad + \sum_{\substack{\text{blocked} \\ w \neq v_0}} d_w \Pr[X'_w \neq Y'_w \mid X, Y] \end{aligned}$$

Due to Lemma 6,

$$\begin{aligned} & \mathbf{E}[\phi_{v_0}(X', Y') \mid X, Y] \\ & = d_{v_0} \Pr[X'_{v_0} \neq Y'_{v_0} \mid X, Y] \\ & \leq d_{v_0} \left[1 - \left(1 - \frac{\Delta}{q}\right) \left(1 - \frac{2}{q}\right)^\Delta \left(1 - \frac{1}{q-2}\right)^{b_{v_0}} \right]. \end{aligned} \tag{21}$$

On the other hand, due to Lemma 7 and Lemma 8,

$$\begin{aligned} & \sum_{u \neq v_0} \mathbf{E}[\phi_u(X', Y') \mid X, Y] \\ & \leq \sum_{\substack{\text{unblocked} \\ u \neq v_0}} \left(\frac{d_u}{q} \left(1 - \frac{2}{q}\right)^{d_u-1} \left[2 - \left(1 - \frac{1}{q-2}\right)^{b_u} \right] \right) \\ & \quad \times \sum_{\substack{\text{unblocked SSAW} \\ \mathcal{P} \text{ from } v_0 \text{ to } u}} \left(\frac{2}{q}\right)^{\ell(\mathcal{P})-1} \\ & \quad + \sum_{\substack{\text{blocked} \\ u \neq v_0}} \frac{d_u}{q} \left(1 - \frac{2}{q}\right)^{d_u-1} \sum_{\substack{\text{SSAW } \mathcal{P} \text{ from } v_0 \text{ to } u \\ \text{with only } u \text{ blocked}}} \left(\frac{2}{q}\right)^{\ell(\mathcal{P})-1} \\ & \leq \sum_{\substack{\text{unblocked} \\ u \neq v_0}} \left(\frac{\Delta}{q} \left(1 - \frac{2}{q}\right)^{\Delta-1} \left[2 - \left(1 - \frac{1}{q-2}\right)^{b_u} \right] \right) \\ & \quad \times \sum_{\substack{\text{unblocked SSAW} \\ \mathcal{P} \text{ from } v_0 \text{ to } u}} \left(\frac{2}{q}\right)^{\ell(\mathcal{P})-1} \\ & \quad + \sum_{\substack{\text{blocked} \\ u \neq v_0}} \frac{\Delta}{q} \left(1 - \frac{2}{q}\right)^{\Delta-1} \sum_{\substack{\text{SSAW } \mathcal{P} \text{ from } v_0 \text{ to } u \\ \text{with only } u \text{ blocked}}} \left(\frac{2}{q}\right)^{\ell(\mathcal{P})-1} \end{aligned} \tag{22}$$

$$\leq \sum_{\substack{\mathcal{P} \text{ from } v_0 \\ \text{to any } u \neq v_0}} \phi_{\mathcal{P}}, \tag{23}$$

where the inequality (22) is due to the monotonicity stated in Lemma 5, and the last sum in (23) enumerates all the walks $\mathcal{P} = (v_0, v_1, \dots, v_\ell)$ from v_0 . And for such walk \mathcal{P} , the quantity $\phi_{\mathcal{P}}$ is defined as that $\phi_{\mathcal{P}} = 0$ if \mathcal{P} is not a strongly self-avoiding walk (SSAW), and for a SSAW $\mathcal{P} = (v_0, v_1, \dots, v_\ell)$ from v_0 to any $v_\ell = u$:

$$\phi_{\mathcal{P}} = \begin{cases} \frac{\Delta}{q} \left(1 - \frac{2}{q}\right)^{\Delta-1} \left[2 - \left(1 - \frac{1}{q-2}\right)^{b_u} \right] \left(\frac{2}{q}\right)^{\ell-1} & \text{(I)} \\ \frac{\Delta}{q} \left(1 - \frac{2}{q}\right)^{\Delta-1} \left(\frac{2}{q}\right)^{\ell-1} & \text{(II)} \\ 0 & \text{(III)} \end{cases}$$

- I : if all v_1, \dots, v_ℓ are unblocked;
- II : if all $v_1, \dots, v_{\ell-1}$ are unblocked and $v_\ell = u$ is blocked;
- III : otherwise.

It is easy to verify the inequality (23) with this definition of $\phi_{\mathcal{P}}$.

Given any walk $\mathcal{P} = (v_0, v_1, \dots, v_\ell)$ from v_0 such that all v_1, \dots, v_ℓ are unblocked, we further define that

$$\Phi_{\mathcal{P}} = \left(\frac{q}{2}\right)^{\ell-1} \sum_{\mathcal{P}' \text{ extends } \mathcal{P}} \phi_{\mathcal{P}'}, \tag{24}$$

where the sum enumerates all walks (not necessarily strongly self-avoiding) $\mathcal{P}' = (v_0, v_1, \dots, v_\ell, v_{\ell+1}, \dots)$ with \mathcal{P} as its prefix, including \mathcal{P} itself.

Then by the inequality (23) the expected distance except for v_0 can be expressed as:

$$\begin{aligned} & \sum_{u \neq v_0} \mathbf{E}[\phi_u(X', Y') \mid X, Y] \\ & \leq \sum_{\substack{\mathcal{P} \text{ from } v_0 \\ \text{to any } u \neq v_0}} \phi_{\mathcal{P}} \\ & = \sum_{u \in \Gamma(v_0) \setminus \Gamma^B(v_0)} \Phi_{(v_0, u)} + \sum_{u \in \Gamma^B(v_0)} \phi_{(v_0, u)} \\ & = \sum_{u \in \Gamma(v_0) \setminus \Gamma^B(v_0)} \Phi_{(v_0, u)} + \frac{\Delta b_{v_0}}{q} \left(1 - \frac{2}{q}\right)^{\Delta-1}. \end{aligned} \tag{25}$$

Here each (v_0, u) is a path (of length 1) from v_0 to its neighbor u .

And more importantly, for $\Phi_{\mathcal{P}}$ we have the following recurrence. For any walk $\mathcal{P} = (v_0, v_1, \dots, v_\ell)$ from v_0 through unblocked vertices $v_1, \dots, v_\ell = u$, if \mathcal{P} is not strongly self-avoiding then $\Phi_{\mathcal{P}} = 0$; and if otherwise \mathcal{P} is strongly self-avoiding, then the following recurrence follows directly from the definition (24) of $\Phi_{\mathcal{P}}$:

$$\begin{aligned} \Phi_{\mathcal{P}} &= \left(\frac{q}{2}\right)^{\ell-1} \phi_{\mathcal{P}} + \left(\frac{q}{2}\right)^{\ell-1} \sum_{w \in \Gamma^B(u)} \phi_{(\mathcal{P}, w)} \\ &\quad + \frac{2}{q} \sum_{\substack{\text{unblocked } w \in \Gamma(u) \\ w \neq v_{\ell-1}}} \Phi_{(\mathcal{P}, w)} \\ &\leq \frac{\Delta}{q} \left(1 - \frac{2}{q}\right)^{\Delta-1} \left[2 - \left(1 - \frac{1}{q-2}\right)^{b_u} + \frac{2b_u}{q}\right] \\ &\quad + \frac{2}{q} \sum_{\substack{\text{unblocked } w \in \Gamma(u) \\ w \neq v_{\ell-1}}} \Phi_{(\mathcal{P}, w)}, \end{aligned} \tag{26}$$

where (\mathcal{P}, w) denotes the walk $\mathcal{P}' = (v_0, v_1, \dots, v_{\ell}, w)$ that extends \mathcal{P} .

The following lemma essentially states that $\Phi_{\mathcal{P}}$ is maximized when the number of blocked neighbors $b_u = 0$ and then the value of $\Phi_{\mathcal{P}}$ is upper bounded by the fixpoint for this recurrence.

Lemma 9 *If $3\Delta < q \leq 3.7\Delta + 3$ and $\Delta \geq 5$, then for any walk $\mathcal{P} = (v_0, v_1, \dots, v_{\ell})$ from v_0 such that all v_1, \dots, v_{ℓ} are unblocked, it holds that*

$$\Phi_{\mathcal{P}} \leq \frac{\Delta}{q - 2\Delta + 2} \left(1 - \frac{2}{q}\right)^{\Delta-1}.$$

Proof We prove by induction on the length of the walk. Let $\mathcal{P} = (v_0, v_1, \dots, v_{\ell})$ be a walk from v_0 such that all v_1, \dots, v_{ℓ} are unblocked and $v_{\ell} = u$. When ℓ is longer than the longest strongly self-avoiding walk among unblocked v_1, \dots, v_{ℓ} , then \mathcal{P} is not a SSAW and thus $\Phi_{\mathcal{P}} = 0$.

Assume that the lemma holds for all unblocked walks longer than ℓ . Then due to the recurrence (26),

$$\begin{aligned} \Phi_{\mathcal{P}} &\leq \frac{\Delta}{q} \left(1 - \frac{2}{q}\right)^{\Delta-1} \left[2 - \left(1 - \frac{1}{q-2}\right)^{b_u} + \frac{2b_u}{q}\right] \\ &\quad + \frac{2}{q} \sum_{\substack{\text{unblocked } w \in \Gamma(u) \\ w \neq v_{\ell-1}}} \Phi_{(\mathcal{P}, w)} \\ \text{(I.H.)} &\leq \frac{\Delta}{q} \left(1 - \frac{2}{q}\right)^{\Delta-1} \left[2 - \left(1 - \frac{1}{q-2}\right)^{b_u} + \frac{2b_u}{q}\right] \\ &\quad + \frac{2(\Delta - b_u - 1)\Delta}{q(q - 2\Delta + 2)} \left(1 - \frac{2}{q}\right)^{\Delta-1} \\ &= \left[1 - \left(1 - \frac{1}{q-2}\right)^{b_u} - \frac{4\Delta - 4}{q(q - 2\Delta + 2)} \cdot b_u\right] \\ &\quad + \frac{\Delta}{q - 2\Delta + 2} \cdot \frac{q}{\Delta} \left[\frac{\Delta}{q} \left(1 - \frac{2}{q}\right)^{\Delta-1}\right], \end{aligned}$$

which is bounded from above by $\frac{\Delta}{q-2\Delta+2} \left(1 - \frac{2}{q}\right)^{\Delta-1}$ if

$$\left(1 - \frac{1}{q-2}\right)^{b_u} + \frac{4\Delta - 4}{q(q - 2\Delta + 2)} \cdot b_u \geq 1.$$

The inequality holds trivially when $b_u = 0$. It is then sufficient to prove that LHS is monotone on integer $b_u \geq 0$:

Denoted $f(x) = \left(1 - \frac{1}{q-2}\right)^x + \frac{4\Delta-4}{q(q-2\Delta+2)} \cdot x$,

$$\begin{aligned} f(b_u + 1) - f(b_u) &= \frac{4\Delta - 4}{q(q - 2\Delta + 2)} - \left(1 - \frac{1}{q-2}\right)^{b_u} \left(\frac{1}{q-2}\right) \\ \text{(since } b_u \geq 0) &\geq \frac{4\Delta - 4}{q(q - 2\Delta + 2)} - \frac{1}{q-2}, \end{aligned}$$

which is nonnegative for $3\Delta - 3 - \sqrt{9\Delta^2 - 26\Delta + 17} \leq q \leq 3\Delta - 3 + \sqrt{9\Delta^2 - 26\Delta + 17}$. In particular this holds when $3\Delta < q \leq 3.7\Delta + 3$ and $\Delta \geq 5$. This completes the induction. \square

Proof of Lemma 4: Combine (21) and (25), with Lemma 9, we obtain

$$\begin{aligned} \mathbf{E}[\Phi(X', Y') \mid X, Y] &= \sum_{u \in V} \mathbf{E}[\phi_u(X', Y') \mid X, Y] \\ &\leq d_{v_0} \left[1 - \left(1 - \frac{\Delta}{q}\right) \left(1 - \frac{2}{q}\right)^{\Delta} \left(1 - \frac{1}{q-2}\right)^{b_{v_0}}\right] \\ &\quad + \frac{\Delta(d_{v_0} - b_{v_0})}{q - 2\Delta + 2} \left(1 - \frac{2}{q}\right)^{\Delta-1} + \frac{\Delta b_{v_0}}{q} \left(1 - \frac{2}{q}\right)^{\Delta-1}. \end{aligned} \tag{27}$$

We need the following technical inequality:

$$\begin{aligned} \left(1 - \frac{\Delta}{q}\right) &\leq \left(1 - \frac{\Delta}{q}\right) \left(1 - \frac{1}{q-2}\right)^{b_{v_0}} \\ &\quad + \frac{2\Delta - 2}{(q-2)(q-2\Delta+2)} b_{v_0} \end{aligned} \tag{28}$$

The equality holds trivially when $b_{v_0} = 0$. It is then sufficient to verify that the RHS is monotone on integer $b_{v_0} \geq 0$. We denote $g(x) = \left(1 - \frac{\Delta}{q}\right) \left(1 - \frac{1}{q-2}\right)^x + \frac{2\Delta-2}{(q-2)(q-2\Delta+2)} x$, and

$$\begin{aligned} g(b_{v_0} + 1) - g(b_{v_0}) &= \frac{q}{(q - 2\Delta + 2)(q - 2)} - \frac{1}{q - 2} \\ &\quad - \left(1 - \frac{\Delta}{q}\right) \left(1 - \frac{1}{q-2}\right)^{b_{v_0}} \frac{1}{q-2} \\ &\geq \frac{q}{(q - 2\Delta + 2)(q - 2)} - \frac{1}{q - 2} - \frac{q - \Delta}{q(q - 2)}, \end{aligned}$$

which is nonnegative if $\frac{q}{(q-2\Delta+2)} \geq 1 + \frac{q-\Delta}{q}$. This easily holds for $\frac{1}{2}(5\Delta - 4 - \sqrt{17\Delta^2 - 32\Delta + 16}) \leq q \leq \frac{1}{2}(5\Delta -$

$4 + \sqrt{17\Delta^2 - 32\Delta + 16}$). In particular, it holds as long as $\Delta \leq q \leq 3.7\Delta + 3$ and $\Delta \geq 9$.

With the inequality (28), the RHS in (27) is maximized when $b_0 = 0$ and hence

$$\begin{aligned} & \mathbb{E}[\Phi(X', Y') \mid X, Y] \\ & \leq d_{v_0} \left[1 - \left(1 - \frac{\Delta}{q}\right) \left(1 - \frac{2}{q}\right)^\Delta + \frac{\Delta}{q - 2\Delta + 2} \left(1 - \frac{2}{q}\right)^{\Delta-1} \right]. \end{aligned}$$

Recall that $\Phi(X, Y) = d_{v_0}$. The path coupling condition (20) holds when there is a constant $\delta > 0$ such that

$$\left(1 - \frac{\Delta}{q}\right) \left(1 - \frac{2}{q}\right)^\Delta - \frac{\Delta}{q - 2\Delta + 2} \left(1 - \frac{2}{q}\right)^{\Delta-1} \geq \delta. \quad (29)$$

For $q = \alpha^* \Delta$ and $\Delta \rightarrow \infty$, then the LHS becomes $e^{-2/\alpha^*} \left(1 - \frac{1}{\alpha^*} - \frac{1}{\alpha^*-2}\right)$, which equals 0 if $\alpha^* = 2 + \sqrt{2}$.

Furthermore, for $q \geq \alpha \Delta$, the LHS become:

$$\begin{aligned} & \left(1 - \frac{\Delta}{q}\right) \left(1 - \frac{2}{q}\right)^\Delta - \frac{\Delta}{q - 2\Delta + 2} \left(1 - \frac{2}{q}\right)^{\Delta-1} \\ & \geq \left(1 - \frac{2}{q}\right)^\Delta \left(1 - \frac{\Delta}{q} - \frac{\Delta}{q - 2\Delta}\right) \\ & \geq \left(1 - \frac{2}{\alpha \Delta}\right)^\Delta \left(1 - \frac{1}{\alpha} - \frac{1}{\alpha - 2}\right) \\ & \geq \left(1 - \frac{2}{\alpha}\right) \left(1 - \frac{1}{\alpha} - \frac{1}{\alpha - 2}\right) \end{aligned}$$

which is a positive constant independent of Δ when $\alpha > \alpha^* = 2 + \sqrt{2}$.

Altogether, by the path coupling Lemma 2, if $\alpha \Delta \leq q \leq 3.7\Delta + 3$ for a constant $\alpha > 2 + \sqrt{2}$ and $\Delta \geq 9$, then the mixing rate is bounded by $\tau(\epsilon) = O(\log(\frac{n}{\epsilon}))$. \square

5 Lower bounds

In this section, we show lower bounds for local sampling. Let $G(V, E)$ be a network, and \mathcal{I} an instance of MRF or weighted local CSP defined on graph G . For example, $\mathcal{I} = (G, [q], \mathbf{A}, \mathbf{b})$ for a MRF with edge activities $\mathbf{A} = \{A_e\}_{e \in E}$ and vertex activities $\mathbf{b} = \{b_v\}_{v \in V}$.

We assume that each vertex $v \in V$ may access to an independent random variable Ψ_v as its source of randomness. Then a t -round protocol specifies a family of functions $\Pi_{v, \mathcal{I}}$, such that for each vertex $v \in V$, the output X_v is produced as

$$X_v = \Pi_{v, \mathcal{I}}(\Psi_u, u \in B_t(v)),$$

where $B_t(v) = \{u \in V \mid \text{dist}(u, v) \leq t\}$ represents the t -ball centered at v . Let μ_{out} denote the distribution of the

output random vector $\mathbf{X} = (X_v)_{v \in V}$. The goal is to have $d_{\text{TV}}(\mu_{\text{out}}, \mu) \leq \epsilon$, where $\mu = \mu_{\mathcal{I}}$ is the Gibbs distribution defined by the MRF instance \mathcal{I} .

Note that in above we allow the protocol $\Pi_{v, \mathcal{I}}$ executed at each vertex $v \in V$ to be aware of the instance \mathcal{I} of the MRF. This is much stronger than the original LOCAL model. In fact, the only locality property we are using to prove our lower bounds is that for any $\mathbf{X} = (X_v)_{v \in V}$ returned by a t -round protocol:

$$\begin{aligned} & \forall u, v \in V : \\ & \text{dist}(u, v) > 2t \implies X_u \text{ and } X_v \text{ are independent.} \end{aligned} \quad (30)$$

The lower bounds implied by this property is due to the locality of randomness.

For many natural MRFs, the Gibbs distribution μ exhibits the following exponential correlations: There exist constants $\delta, \eta > 0$ such that for a path P of length n , any vertices u, v from the path, there are two spin states $\sigma_u, \sigma'_u \in [q]$ such that $\mu_u(\sigma_u) \geq \delta, \mu_u(\sigma'_u) \geq \delta$ for the marginal distribution μ_u induced by μ at vertex u and

$$d_{\text{TV}}(\mu_v(\cdot \mid \sigma_u), \mu_v(\cdot \mid \sigma'_u)) \geq \eta^{\text{dist}(u, v)}. \quad (31)$$

This exponential correlation property is satisfied by many MRFs, in particular, the proper q -colorings for any constant q . For MRFs having this property, for any $\epsilon > \exp(-o(n))$, vertex pairs (u, v) with sufficiently small $\text{dist}(u, v) = \Omega(\log \frac{1}{\epsilon})$ will contribute at least an ϵ total variation distance between Gibbs (σ_u, σ_v) and any independent (X_u, X_v) . And due to (30), this gives an $\Omega(\log \frac{1}{\epsilon})$ lower bound for local sampling from any MRF satisfying (31), where ϵ is the total variation distance.

We then show that the $\Omega(\log n)$ lower bound holds even for a constant total variation distance ϵ . A similar $\Omega(\log n)$ lower bound for sampling independent sets is proved independently in [33]. Altogether it shows that the $O(\log(\frac{n}{\epsilon}))$ upper bound in Theorem 2 is optimal.

Theorem 7 *Let $q \geq 3$ be a constant and $\epsilon < \frac{1}{3}$. Any t -round protocol that samples uniform proper q -coloring in a path within total variation distance ϵ must have $t = \Omega(\log n)$.*

Proof We actually prove the lower bound for all MRFs satisfying a stronger exponential correlation property stated as follows: There exist constants $\delta, \eta > 0$ such that for a path P of length n , for any non-adjacent vertices x, u, v, y in the path from left to right, any spin states $\sigma_x, \sigma_y \in [q]$, there exist two spin states $\sigma_u, \sigma'_u \in [q]$ such that $\mu_u(\sigma_u \mid \sigma_x, \sigma_y) \geq \delta, \mu_u(\sigma'_u \mid \sigma_x, \sigma_y) \geq \delta$ and

$$d_{\text{TV}}(\mu_v(\cdot \mid \sigma_u, \sigma_x, \sigma_y), \mu_v(\cdot \mid \sigma'_u, \sigma_x, \sigma_y)) \geq \eta^{\text{dist}(u, v)}. \quad (32)$$

It can be verified by a simple recursion for marginal probabilities in paths [45] that this property as well as the weaker correlation property (31) hold for uniform proper q -colorings in paths for any constant $q \geq 3$.

Let $P = (w_0, w_1, \dots, w_{n-1})$ be a path of n vertices. For $i = 0, 1, \dots, m$ where $m = \lfloor \frac{n-1}{3(2t+1)} \rfloor$, we denote $x_i = w_{3(2t+1)i}$; and for $i = 0, 1, \dots, m-1$, denote $u_i = w_{3(2t+1)i+2t+1}$, and $v_i = w_{3(2t+1)i+2(2t+1)}$. We denote $F = \{x_i \mid 0 \leq i \leq m\}$ and $U = \{u_i, v_i \mid 0 \leq i \leq m-1\}$, and let $C = F \cup U$. We call the vertices in C the *centers*, and the vertices in F and U the *fixed* and *unfixed* centers respectively. Note that the pairs (u_i, v_i) of consecutive unfixed centers are separated by the fixed centers x_i 's. Due to the conditional independence of MRF, conditioning on any particular configuration $\sigma_F \in [q]^F$ of fixed centers, for a $\sigma \in [q]^P$ sampled from the Gibbs distribution μ consistent with σ_F over F , the pairs $(\sigma_{u_i}, \sigma_{v_i})$ are mutually independent of each other. For the followings we assume that we are conditioning on an arbitrarily fixed $\sigma_F \in [q]^F$.

Let X_{u_i} and X_{v_i} be the respective output of u_i and v_i in a t -round protocol. Due to the observation of (30), X_{u_i} and X_{v_i} are mutually independent. According to the exponential correlation of (32), by choosing a suitably small $t = O(\log n)$, the total variation distance between $(\sigma_{u_i}, \sigma_{v_i})$ and (X_{u_i}, X_{v_i}) is at least $\exp(-\Omega(t)) = n^{-\frac{1}{2}}$.

We denote $\mathcal{X}_i = (X_{u_i}, X_{v_i})$ and $\mathcal{Y}_i = (\sigma_{u_i}, \sigma_{v_i})$, and consider the random vector $\mathcal{X} = (\mathcal{X}_i)_{0 \leq i \leq m-1}$ and $\mathcal{Y} = (\mathcal{Y}_i)_{0 \leq i \leq m-1}$ where \mathcal{Y} is sampled conditioning on an arbitrarily fixed $\sigma_F \in [q]^F$. As we argued above, both $\mathcal{X} = (\mathcal{X}_i)$ and $\mathcal{Y} = (\mathcal{Y}_i)$ are vectors of mutually independent variables, and $d_{TV}(\mathcal{X}_i, \mathcal{Y}_i) \geq n^{-\frac{1}{2}}$. Therefore, for any coupling of \mathcal{X} and \mathcal{Y} , we have

$$\begin{aligned} \Pr[\mathcal{X} \neq \mathcal{Y}] &= 1 - \prod_{i=0}^{m-1} \Pr[\mathcal{X}_i = \mathcal{Y}_i \mid \forall j < i, \mathcal{X}_j = \mathcal{Y}_j] \\ &\geq 1 - \left(1 - n^{-1/2}\right)^m. \end{aligned} \tag{33}$$

Note that in an arbitrary coupling $(\mathcal{X}, \mathcal{Y})$, the pairs $(\mathcal{X}_i, \mathcal{Y}_i)$ are not necessarily mutually independent of each other even though \mathcal{X}_i 's (and \mathcal{Y}_i 's) are mutually independent in \mathcal{X} (and in \mathcal{Y}). Nevertheless, conditioning on $(\mathcal{X}_j, \mathcal{Y}_j)$ for $j < i$ will only affect the joint distribution of $(\mathcal{X}_i, \mathcal{Y}_i)$ but not the marginal distributions of \mathcal{X}_i and \mathcal{Y}_i because of the mutual independence between \mathcal{X}_i 's (and between \mathcal{Y}_i 's). And by the coupling lemma, we have $\Pr[\mathcal{X}_i = \mathcal{Y}_i] \leq 1 - d_{TV}(\mathcal{X}_i, \mathcal{Y}_i) \leq 1 - n^{-\frac{1}{2}}$ for any coupling of $(\mathcal{X}_i, \mathcal{Y}_i)$. The inequality (33) follows.

Since (33) holds for any coupling $(\mathcal{X}, \mathcal{Y})$, applying the coupling lemma again, we obtain that

$$d_{TV}(\mathcal{X}, \mathcal{Y}) \geq 1 - \left(1 - n^{-1/2}\right)^m = 1 - o(1), \tag{34}$$

when $t = O(\log n)$ and $m = \Omega(n/\log n)$.

Recall that the above \mathcal{Y} is sampled conditioning on an arbitrary configuration $\sigma_F \in [q]^F$ of fixed centers. Now we consider a $\sigma \in [q]^P$ sampled from the Gibbs distribution μ on the path P and its restrictions σ_F, σ_U and σ_C on $F = \{x_i\}$, $U = \{u_i, v_i\}$ and $C = F \cup U$. Also let X be the vector of values returned by the vertices in P in a t -round protocol, and X_F, X_U and X_C its restrictions on the respective sets of centers. The theorem follows if we can show that $d_{TV}(X, \sigma) > \frac{1}{3}$ for our choice of $t = O(\log n)$. By definition of the total variation distance, we have:

$$\begin{aligned} d_{TV}(X, \sigma) &\geq d_{TV}(X_C, \sigma_C) \\ &= \frac{1}{2} \sum_{\sigma_F \in [q]^F} \sum_{\sigma_U \in [q]^U} \left(|\mu(\sigma_F, \sigma_U) - \Pr[X_F = \sigma_F \wedge X_U = \sigma_U]| \right) \\ &= \frac{1}{2} \sum_{\sigma_F \in [q]^F} \sum_{\sigma_U \in [q]^U} \left(|\mu(\sigma_F)\mu(\sigma_U \mid \sigma_F) - \Pr[X_F = \sigma_F] \Pr[X_U = \sigma_U]| \right) \\ &\geq \sum_{\sigma_F \in [q]^F} \mu(\sigma_F) \cdot \frac{1}{2} \sum_{\sigma_U \in [q]^U} |\mu(\sigma_U \mid \sigma_F) - \Pr[X_U = \sigma_U]| \\ &\quad - \frac{1}{2} \sum_{\sigma_F \in [q]^F} |\mu(\sigma_F) - \Pr[X_F = \sigma_F]|. \end{aligned} \tag{35}$$

Note that

$$\begin{aligned} d_{TV}(X, \sigma) &\geq d_{TV}(X_F, \sigma_F) \\ &= \frac{1}{2} \sum_{\sigma_F \in [q]^F} |\mu(\sigma_F) - \Pr[X_F = \sigma_F]|. \end{aligned}$$

If this quantity is greater than $1/3$, then we already have $d_{TV}(X, \sigma) > 1/3$ and the lower bound is proved. If otherwise, we suppose that

$$\frac{1}{2} \sum_{\sigma_F \in [q]^F} |\mu(\sigma_F) - \Pr[X_F = \sigma_F]| \leq \frac{1}{3}.$$

Observe that for any $\sigma_F \in [q]^F$, we have

$$\begin{aligned} \frac{1}{2} \sum_{\sigma_U \in [q]^U} |\mu(\sigma_U \mid \sigma_F) - \Pr[X_U = \sigma_U]| &= d_{TV}(X, \mathcal{Y}) \\ &\geq 1 - o(1), \end{aligned}$$

where $\mathcal{Y} = (\mathcal{Y}_i = (\sigma_{u_i}, \sigma_{v_i}))_{0 \leq i \leq m-1}$ is sampled conditioning on σ_F and the inequality is due to (34).

Therefore, the total variation distance in (35) can be further bounded as

$$d_{TV}(X, \sigma) \geq \sum_{\sigma_F \in \{q\}^F} \mu(\sigma_F)(1 - o(1)) - \frac{1}{3} = 1 - o(1) - \frac{1}{3} > \frac{1}{3}.$$

□

Next, we state a strong $\Omega(\text{diam})$ lower bound for sampling with long-range correlations.

5.1 An $\Omega(\text{diam})$ lower bound in the non-uniqueness regime

We consider the weighted independent sets of graphs, the *hardcore model*. Given a graph $G(V, E)$ and a *fugacity* parameter $\lambda > 0$, each configuration σ in

$$IS(G) = \left\{ \sigma \in \{0, 1\}^V : \forall (u, v) \in E, \sigma_u \sigma_v = 0 \right\}$$

indicates an independent set I in G and is assigned a weight $w(\sigma) = \lambda^{|I|}$. The Gibbs distribution $\mu = \mu_G$ is defined over all independent sets in G proportional to their weights. As discussed in Sect. 2.2, the model is an MRF.

The hardcore model on graphs with maximum degree Δ undergoes a computational phase transition at the *uniqueness threshold* $\lambda_c(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^\Delta}$, such that sampling from the Gibbs distribution can be done in polynomial time in the *uniqueness regime* $\lambda < \lambda_c$ [20,60] and is intractable unless NP=RP in the *non-uniqueness regime* $\lambda > \lambda_c$ [8,28,55,56].

The following theorem states an $\Omega(\text{diam})$ lower bound for sampling from the hardcore model in the non-uniqueness regime. In particular when $\lambda = 1$ the model represents the uniform independent sets and the non-uniqueness $\lambda > \lambda_c(\Delta)$ holds when $\Delta \geq 6$, which gives us Theorem 3.

Theorem 8 *Let $\Delta \geq 3$ and $\lambda > \lambda_c(\Delta)$. Let $\epsilon > 0$ be a sufficiently small constant. For all $N > 0$ there exists a graph \mathcal{G} on $\Theta(N)$ vertices with maximum degree Δ and diameter $\text{diam}(\mathcal{G}) = \Omega(N^{1/11})$ such that for the hardcore model on \mathcal{G} with fugacity λ , any t -round protocol that samples within total variation distance ϵ from the Gibbs distribution $\mu = \mu_{\mathcal{G}}$ must have $t = \Omega(\text{diam}(\mathcal{G}))$.*

We follow the approaches in [8,27,28,55,56] for the computational phase transition. The network $\mathcal{G} = H^G$ is constructed by lifting a graph H with a gadget G , such that sampling from the hardcore model on H^G with $\lambda > \lambda_c(\Delta)$ effectively samples a maximum cut in H . We choose H to be an even cycle, in which the maximum cut imposes a long-range correlation among vertices. And to sample with such a long-range correlation, the sampling algorithm must not be local.

Unlike the results of [8,27,28,55,56] which are for computational complexity of approximate counting, here we prove unconditional lower bounds for sampling in the LOCAL model. Our lower bound is due to the long-range correlations in the random max-cut rather than the computational complexity of optimization. Technical-wise, this means that in addition to show that a max-cut in H is sampled, we also need that the sampled max-cut is distributed almost uniformly.

5.1.1 The random graph gadget

We now describe the random graph gadget which is essential to the hardness of sampling. The gadget is constructed in two steps. For positive integers n, r and Δ , we first describe the construction of the random bipartite (multi)graph \mathcal{G}'_n :

- Let V^+ and V^- be two vertex sets with $|V^+| = |V^-| = n + r$, such that $V^\pm = U^\pm \uplus W^\pm$ where $|U^\pm| = n$ and $|W^\pm| = r$. Let $V = V^+ \cup V^-$, $W = W^+ \cup W^-$ and $U = U^+ \cup U^-$.
- Uniformly and independently sample $\Delta - 1$ perfect matchings between V^+ and V^- and then uniformly and independently sample a perfect matching between U^+ and U^- . The union of all these matchings gives us the random bipartite (multi)graph \mathcal{G}'_n , in which every vertex in U has degree Δ and every vertex in W has degree $\Delta - 1$.

Now we describe the second part of the construction. Let $0 < \theta < \psi < 1/8$ be constants. Let $r' := (\Delta - 1)^{\lceil \theta \log_{\Delta-1} n \rceil + 2 \lfloor \frac{\psi}{2} \log_{\Delta-1} n \rfloor}$. Note that $r' = o(n^{1/4})$. First, we sample G from the distribution $\mathcal{G}'_{r'}$. Next, attach k disjoint $(\Delta - 1)$ -ary trees of even depth l (with $k = (\Delta - 1)^{\lceil \theta \log_{\Delta-1} n \rceil}$ and $l = 2 \lfloor \frac{\psi}{2} \log_{\Delta-1} n \rfloor$) to W^\pm , such that every vertex in W is a leaf of exactly one tree and the trees do not share common vertices with the bipartite graph G , apart from the vertices in W . Let T^\pm denote the roots of those trees ($|T^+| = |T^-| = k$), called “terminals”. We denote the family of graphs that can be constructed this way by $\tilde{\mathcal{G}}(k, n, \Delta)$. Note that our construction is still bipartite with size $\Theta(n)$ and the terminals in T^+ and T^- belongs to distinct partitions of the bipartite graph.

The *phase* of a configuration σ , denoted as $Y(\sigma)$, is defined as

$$Y(\sigma) := \begin{cases} + & \text{if } \sum_{v \in U^+} \sigma_v \geq \sum_{v \in U^-} \sigma_v, \\ - & \text{if } \sum_{v \in U^+} \sigma_v < \sum_{v \in U^-} \sigma_v. \end{cases}$$

It is easy to verify that the random bipartite graph \mathcal{G}'_n in the first step is an expander with high probability. The following proposition was proved in [8].

Proposition 4 (Lemma 8 & Lemma 9 in [8]) *If $\lambda > \lambda_c(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^\Delta}$ then there exist two constants $0 < q^- < q^+ < 1$ such that the followings hold. Let Q_T^\pm denote the product measure on configurations in $\{0, 1\}^T$ so that the spin states are i.i.d. Bernoulli with probability q^\pm on T^+ and q^\mp on T^- , that is:*

$$Q_T^\pm(\sigma_T) = (q^\pm)^{\sum_{v \in T^+} \sigma_v} (1 - q^\pm)^{|T^+| - \sum_{v \in T^+} \sigma_v} \cdot (q^\mp)^{\sum_{v \in T^-} \sigma_v} (1 - q^\mp)^{|T^-| - \sum_{v \in T^-} \sigma_v}.$$

For any $\delta > 0$, there exists sufficiently large constant $N_0(\delta)$ such that for all $n > N_0(\delta)$ the followings hold altogether with positive probability for $G \sim \tilde{G}(k, n, \Delta)$:

- (expander) G is connected with $\text{diam}(G) = O(\log n)$;
- (balanced phases) $\Pr_G [Y(\sigma) = \pm] \in [(1 - \delta)/2, (1 + \delta)/2]$;
- (phase-correlated almost independence) $\forall \tau_T \in \{0, 1\}^T$, $\Pr_G [\sigma_T = \tau_T \mid Y(\sigma) = \pm] / Q_T^\pm(\tau_T) \in [1 - \delta, 1 + \delta]$;

where \Pr_G is the probability law for σ sampled from μ_G .

By the probabilistic method, there exists a G satisfying the above conditions.

5.1.2 Reduction from max-cut

Let H be a cycle with m vertices where $m > 0$ is an even integer. Fix constants $\theta = \psi = 1/9$ and let $G \in \tilde{G}(2k, n, \Delta)$, with $k = \Theta(m^{10/9})$ and $n = \Theta(k^{1/\theta}) = \Theta(m^{10})$, be the graph that satisfies the conditions in Proposition 4.

- For each vertex $x \in H$ let G_x be a copy of G . We denote by T_x^\pm the respective set of $2k$ terminals in G_x . Let \widehat{H}^G be the disconnected copies of the G_x , $x \in H$.
- For every edge $(x, y) \in H$, add k edges between T_x^+ and T_y^+ and similarly add k edges between T_x^- and T_y^- . This can be done in such a way that the resulting (multi)graph H^G is Δ -regular.

Definition 4 For each $x \in H$, we write $Y_x = Y_x(\sigma)$ for the phase of a configuration σ on G_x . Let $\mathcal{Y} = (Y_x)_{x \in H} \in \{+, -\}^{V(H)}$. Given the phase $\mathcal{Y}' \in \{+, -\}^{V(H)}$, we define:

$$Z_{H^G}(\mathcal{Y}') = \sum_{\sigma \in \text{IS}(H^G)} \lambda^{|\sigma|} \mathbf{1}\{\mathcal{Y}(\sigma) = \mathcal{Y}'\},$$

where

$$\text{IS}(H^G) = \left\{ \sigma \in \{0, 1\}^{V(H^G)} : \forall uv \in E(H^G), \sigma_u \sigma_v = 0 \right\}$$

is the set of all independent sets in H^G . We also use \Pr_{H^G} to represent the probability law for σ sampled from μ_{H^G} .

Note that the cycle H has precisely two maximum cuts. A key property for proving the lower bound is that in the non-uniqueness regime, sampling from the hardcore model on graph H^G corresponds to sampling a maximum cut in H almost uniformly.

Theorem 9 *Let $\lambda > \lambda_c(\Delta)$. Let $\mathcal{Y}_1, \mathcal{Y}_2 \in \{+, -\}^{V(H)}$ correspond respectively to the two maximum cuts in H . It holds that:*

$$\Pr_{H^G} [\mathcal{Y}(\sigma) = \mathcal{Y}_1] = \Pr_{H^G} [\mathcal{Y}(\sigma) = \mathcal{Y}_2] \geq \frac{1}{2} - o(1). \tag{36}$$

The theorem is implied by the following lemma, which is proved by applying a calculation in [55] with the improved gadget property Proposition 4.

Lemma 10 *Let $\mathcal{Y}', \mathcal{Y}'' \in \{+, -\}^{V(H)}$ and $\delta > 0$. Suppose that G satisfies the conditions in Proposition 4. It holds that*

$$\frac{\Pr_{H^G} [\mathcal{Y}(\sigma) = \mathcal{Y}']}{\Pr_{H^G} [\mathcal{Y}(\sigma) = \mathcal{Y}']} \geq \left(\frac{1 - \delta}{1 + \delta} \right)^{2m} (\Theta/\Gamma)^{k[\text{Cut}(\mathcal{Y}') - \text{Cut}(\mathcal{Y}'')]} ,$$

where $\Theta = (1 - q^+ q^-)^2$ and $\Gamma = (1 - (q^+)^2)(1 - (q^-)^2)$; and $\text{Cut}(\mathcal{Y}) = |\{(x, y) \in E(H) : \mathcal{Y}_x \neq \mathcal{Y}_y\}|$ for a $\mathcal{Y} \in \{+, -\}^{V(H)}$.

Proof Since the graph \widehat{H}^G consists of a collection of disconnected copies of G , the distribution of a configuration on \widehat{H}^G is given by the product measure of configurations on the $(G_x)_{x \in H}$. In particular the phases are independent, therefore

$$\begin{aligned} \frac{Z_{\widehat{H}^G}(\mathcal{Y}')}{Z_{\widehat{H}^G}(\mathcal{Y}'')} &= \frac{Z_{\widehat{H}^G}(\mathcal{Y}')/Z_{\widehat{H}^G}}{Z_{\widehat{H}^G}(\mathcal{Y}'')/Z_{\widehat{H}^G}} \\ &= \frac{\Pr_G [Y(\sigma) = +]^{\sum_{x \in H} \mathbf{1}\{Y'_x = +\}} \cdot \Pr_G [Y(\sigma) = -]^{\sum_{x \in H} \mathbf{1}\{Y'_x = -\}}}{\Pr_G [Y(\sigma) = +]^{\sum_{x \in H} \mathbf{1}\{Y''_x = +\}} \cdot \Pr_G [Y(\sigma) = -]^{\sum_{x \in H} \mathbf{1}\{Y''_x = -\}}} \\ &\geq \left(\frac{1 - \delta}{1 + \delta} \right)^m. \end{aligned} \tag{37}$$

Note that the ratio $Z_{H^G}(\mathcal{Y}')/Z_{\widehat{H}^G}(\mathcal{Y}')$ is precisely the probability of a σ sampled from $\mu_{\widehat{H}^G}$ being an independent set in H^G . And due to Proposition 4, conditioning on the phase \mathcal{Y}' the spins of $\sigma_{\cup_{x \in H} T_x}$ are almost independent i.i.d. Bernoulli with probabilities q^+ or q^- depending on the phase, therefore

$$\begin{aligned} \frac{Z_{H^G}(\mathcal{Y}')}{Z_{\widehat{H}^G}(\mathcal{Y}')} &= \Pr_{\widehat{H}^G} [\sigma \text{ is an IS in } H^G \mid \mathcal{Y}(\sigma) = \mathcal{Y}'] \\ &= \Pr_{\widehat{H}^G} [\forall (u, v) \in E(H^G) \setminus E(\widehat{H}^G), \sigma_u \sigma_v \neq 1 \mid \mathcal{Y}(\sigma) = \mathcal{Y}'] \\ &\geq (1 - \delta)^m \sum_{\sigma_{\cup_{x \in H} T_x}} Q_{\sigma_T}(\mathcal{Y}') \\ &= (1 - \delta)^m \Gamma^{k|E(H)|} (\Theta/\Gamma)^{k \text{Cut}(\mathcal{Y}')}, \end{aligned} \tag{38}$$

where

$$\mathcal{Q}_{\sigma_T}(\mathcal{Y}') = \left[\mathbf{1}(\forall(u, v) \in E(H^G) \setminus E(\widehat{H}^G), \sigma_u \sigma_v \neq 1) \times \prod_{x \in H} \mathcal{Q}_{T_x}^{Y'_x}(\sigma_{T_x}) \right].$$

Similarly, we can obtain

$$\frac{Z_{H^G}(\mathcal{Y}'')}{Z_{\widehat{H}^G}(\mathcal{Y}'')} \leq (1 + \delta)^m \Gamma^{k|E(H)|} (\Theta/\Gamma)^{k \text{Cut}(\mathcal{Y}'')}. \tag{39}$$

Combining (37), (38) and (39), we have:

$$\begin{aligned} \frac{\Pr_{H^G} [\mathcal{Y}(\sigma) = \mathcal{Y}']}{\Pr_{H^G} [\mathcal{Y}(\sigma) = \mathcal{Y}'']} &= \frac{Z_{H^G}(\mathcal{Y}')}{Z_{H^G}(\mathcal{Y}'')} \\ &\geq \left(\frac{1 - \delta}{1 + \delta} \right)^m (\Theta/\Gamma)^{k[\text{Cut}(\mathcal{Y}') - \text{Cut}(\mathcal{Y}'')]} \cdot \frac{Z_{\widehat{H}^G}(\mathcal{Y}')}{Z_{\widehat{H}^G}(\mathcal{Y}'')} \\ &\geq \left(\frac{1 - \delta}{1 + \delta} \right)^{2m} (\Theta/\Gamma)^{k[\text{Cut}(\mathcal{Y}') - \text{Cut}(\mathcal{Y}'')]} \end{aligned}$$

□

Proof of Theorem 9: Let $\mathcal{Y}', \mathcal{Y}'' \in \{+, -\}^{V(H)}$ such that $\text{Cut}(\mathcal{Y}') > \text{Cut}(\mathcal{Y}'')$. Let $\delta > 0$, by Lemma 10, we have

$$\frac{\Pr_{H^G} [\mathcal{Y}(\sigma) = \mathcal{Y}']}{\Pr_{H^G} [\mathcal{Y}(\sigma) = \mathcal{Y}'']} \geq \left(\frac{1 - \delta}{1 + \delta} \right)^{2m} (\Theta/\Gamma)^{k[\text{Cut}(\mathcal{Y}') - \text{Cut}(\mathcal{Y}'')]}.$$

Note that for $\lambda > \lambda_c(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^\Delta}$, we have $\Theta > \Gamma$. Thus for $k = \Theta(m^{10/9})$ we have

$$\frac{\Pr_{H^G} [\mathcal{Y}(\sigma) = \mathcal{Y}']}{\Pr_{H^G} [\mathcal{Y}(\sigma) = \mathcal{Y}'']} \geq \left(\frac{1 - \delta}{1 + \delta} \right)^{2m} (\Theta/\Gamma)^k \geq 4^m.$$

Since the size of $\{+, -\}^{V(H)}$ is at most 2^m , it follows that with probability at least $1 - o(1)$ the phases $\mathcal{Y}(\sigma)$ attain a maximum cut in H . Therefore, we only need to prove $Z_{H^G}(\mathcal{Y}_1) = Z_{H^G}(\mathcal{Y}_2)$ for the two maximum cuts \mathcal{Y}_1 and \mathcal{Y}_2 in H . By simple calculation, we have

$$\begin{aligned} Z_{H^G}(\mathcal{Y}_1) &= Z_{\widehat{H}^G}(\mathcal{Y}_1) \cdot \Pr_{\widehat{H}^G} [\sigma \in \text{IS}(H^G) \mid \mathcal{Y}(\sigma) = \mathcal{Y}_1] \\ &= Z_{\widehat{H}^G}(\mathcal{Y}_1) \cdot \Pr_{\widehat{H}^G} [\forall(u, v) \in E(H^G) \setminus E(\widehat{H}^G), \sigma_u \sigma_v \neq 1 \mid \mathcal{Y}(\sigma) = \mathcal{Y}_1] \\ &= Z_{\widehat{H}^G} \cdot \Pr_G [Y = +]^{m/2} \cdot \Pr_G [Y = -]^{m/2} \\ &\cdot \Pr_{\widehat{H}^G} [\forall(u, v) \in E(H^G) \setminus E(\widehat{H}^G), \sigma_u \sigma_v \neq 1 \mid \mathcal{Y}(\sigma) = \mathcal{Y}_1] \end{aligned}$$

and

$$\begin{aligned} Z_{H^G}(\mathcal{Y}_2) &= Z_{\widehat{H}^G}(\mathcal{Y}_2) \cdot \Pr_{\widehat{H}^G} [\sigma \in \text{IS}(H^G) \mid \mathcal{Y}(\sigma) = \mathcal{Y}_2] \\ &= Z_{\widehat{H}^G}(\mathcal{Y}_2) \cdot \Pr_{\widehat{H}^G} [\forall(u, v) \in E(H^G) \setminus E(\widehat{H}^G), \sigma_u \sigma_v \neq 1 \mid \mathcal{Y}(\sigma) = \mathcal{Y}_2] \\ &= Z_{\widehat{H}^G} \cdot \Pr_G [Y = +]^{m/2} \cdot \Pr_G [Y = -]^{m/2} \\ &\cdot \Pr_{\widehat{H}^G} [\forall(u, v) \in E(H^G) \setminus E(\widehat{H}^G), \sigma_u \sigma_v \neq 1 \mid \mathcal{Y}(\sigma) = \mathcal{Y}_2]. \end{aligned}$$

By symmetry of the even-length cycle, it holds that

$$\begin{aligned} \Pr_{\widehat{H}^G} [\forall(u, v) \in E(H^G) \setminus E(\widehat{H}^G), \sigma_u \sigma_v \neq 1 \mid \mathcal{Y}(\sigma) = \mathcal{Y}_1] &= \Pr_{\widehat{H}^G} [\forall(u, v) \in E(H^G) \setminus E(\widehat{H}^G), \sigma_u \sigma_v \neq 1 \mid \mathcal{Y}(\sigma) = \mathcal{Y}_2]. \end{aligned}$$

□

5.1.3 Proof of the $\Omega(\text{diam})$ lower bound

Now we are ready to prove Theorem 8. Let N be sufficiently large. We choose an integer $n = \Theta(N^{10/11})$ and even integer $m = \Theta(N^{1/11})$ such that $m/2$ is odd, so that a gadget G is constructed to satisfy Proposition 4, and the graph $\mathcal{G} = H^G$, where H is a cycle of length m , is constructed as described in Sect. 5.1.2. Note that $\text{diam}(\mathcal{G}) \geq \text{diam}(H) \geq m/2$ and $|V(\mathcal{G})| = \Theta(N)$, therefore $\text{diam}(\mathcal{G}) = \Omega(N^{1/11})$.

Let σ' denote the output of a t -round protocol with $t \leq 0.49 \cdot \text{diam}(\mathcal{G})$ on network \mathcal{G} , whose distribution is denoted as μ_t ; and let σ be sampled from the hardcore Gibbs distribution $\mu = \mu_{\mathcal{G}}$. By contradiction, we assume that $d_{\text{TV}}(\mu_t, \mu) \leq \epsilon$ for sufficiently small constant ϵ .

Let $\mathcal{Y}', \mathcal{Y}'' \in \{+, -\}^{V(H)}$ denote the phases corresponding to the two maximum cuts in the cycle H . Therefore, by Theorem 9, we have

$$\Pr[\mathcal{Y}(\sigma) \in \{\mathcal{Y}', \mathcal{Y}''\}] \geq 1 - o(1).$$

We pick $u, v \in V(\mathcal{G})$ which satisfy that $\text{dist}_{\mathcal{G}}(u, v) = \text{diam}(\mathcal{G})$. Since $\mathcal{G} = H^G$ is constructed by replacing each vertex x in H with G_x which is an identical copy of G , it must hold that $u \in G_x, v \in G_y$ for some vertices x, y in H with $\text{dist}_H(x, y) = m/2$. And since $m/2$ is odd, without loss of generality, we suppose that $Y'_x = +, Y'_y = -$ and $Y''_x = -, Y''_y = +$. Moreover, for all $u' \in G_x, v' \in G_y$, by the triangle inequality we have:

$$\begin{aligned} \text{dist}_{\mathcal{G}}(u, u') + \text{dist}_{\mathcal{G}}(u', v') + \text{dist}_{\mathcal{G}}(v', v) &\geq \text{dist}_{\mathcal{G}}(u, v) = \text{diam}(\mathcal{G}). \end{aligned}$$

Due to Proposition 4, it holds that $\text{diam}(G) = O(\log n)$, thus we have:

$$\text{dist}_{\mathcal{G}}(u', v') \geq \text{diam}(\mathcal{G}) - O(\log n) = (1 - o(1))\text{diam}(\mathcal{G}).$$

For the σ' returned by a t -round protocol where $t \leq 0.49 \cdot \text{diam}(\mathcal{G})$, according to the property (30), the σ'_{G_x} and σ'_{G_y} are independent of each other, thus the phases of G_x and G_y on σ' are independent of each other:

$$\begin{aligned} & \Pr [Y_x(\sigma') = + \mid Y_y(\sigma') = -] \\ &= \Pr [Y_x(\sigma') = + \mid Y_y(\sigma') = +]. \end{aligned} \tag{40}$$

On the other hand, since $d_{\text{TV}}(\sigma', \sigma) \leq \epsilon$, we have

$$\begin{aligned} & \Pr [Y_x(\sigma') = + \mid Y_y(\sigma') = -] \\ &= \frac{\Pr [Y_x(\sigma') = + \wedge Y_y(\sigma') = -]}{\Pr [Y_y(\sigma') = -]} \\ &\geq \frac{\Pr [Y_x(\sigma) = + \wedge Y_y(\sigma) = -] - \epsilon}{\Pr [Y_y(\sigma) = -] + \epsilon} \quad (\text{by } d_{\text{TV}}(\sigma', \sigma) \leq \epsilon) \\ &\geq \frac{\Pr [\mathcal{Y}(\sigma) = \mathcal{Y}'] - \epsilon}{\Pr [Y_y(\sigma) = -] + \epsilon} \\ &\geq \frac{1/2 - o(1) - \epsilon}{\Pr [\mathcal{Y}(\sigma) \neq \mathcal{Y}''] + \epsilon} \geq \frac{1 - 2\epsilon - o(1)}{1 + 2\epsilon + o(1)}, \quad (\text{by Theorem 9}) \end{aligned}$$

and

$$\begin{aligned} & \Pr [Y_x(\sigma') = + \mid Y_y(\sigma') = +] \\ &= \frac{\Pr [Y_x(\sigma') = + \wedge Y_y(\sigma') = +]}{\Pr [Y_y(\sigma') = +]} \\ &\leq \frac{\Pr [Y_x(\sigma) = + \wedge Y_y(\sigma) = +] + \epsilon}{\Pr [Y_y(\sigma) = +] - \epsilon} \quad (\text{by } d_{\text{TV}}(\sigma', \sigma) \leq \epsilon) \\ &\leq \frac{\Pr [\mathcal{Y}(\sigma) \notin \{\mathcal{Y}', \mathcal{Y}''\}] + \epsilon}{\Pr [\mathcal{Y}(\sigma) = \mathcal{Y}''] - \epsilon} \\ &\leq \frac{2\epsilon + o(1)}{1 - 2\epsilon - o(1)}. \quad (\text{by Theorem 9}) \end{aligned}$$

This implies that

$$\Pr [Y_x(\sigma') = + \mid Y_y(\sigma') = +] < \Pr [Y_x(\sigma') = + \mid Y_y(\sigma') = -]$$

by taking ϵ to be a sufficiently small constant, which contradicts the independence given in (40).

6 Conclusion

In this paper, we study the local sampling problem and ask a new question about local computation: *whether a locally definable joint distribution can be sampled locally.*

On the positive side, we give two distributed sampling algorithms *LubyGlauber* and *LocalMetropolis*. *LubyGlauber*

achieves $O(\Delta \log n)$ mixing time under Dobrushin’s condition and *LocalMetropolis* may achieve optimal $O(\log n)$ mixing time under a stronger mixing condition. Thus many locally definable joint distributions can be sampled locally.

On the negative side, we give an $\Omega(\log n)$ lower bound for sampling from a broad class of locally defined joint distributions. Thus the $O(\log n)$ -radius can be considered as the new criteria for being local for distributed sampling algorithms. Furthermore, we give an $\Omega(\text{diam}) = n^{\Omega(1)}$ lower bound for sampling weighted independent sets in the non-uniqueness regime. Since independent set is trivial to construct, this gives a strong separation between local sampling and local construction. The lower bounds hold even if every vertex is aware of the graph structure, which means the hardness for local sampling is due to the discrepancy between the locality of randomness in distributed algorithms and the long-range correlation in the joint distribution from which we want to sample.

Acknowledgements This research is supported by the National Key R&D Program of China 2018YFB1003202 and the National Science Foundation of China under Grant Nos. 61722207 and 61672275. Yitong Yin wants to thank Daniel Štefankovič for the stimulating discussions in the beginning of this project. He also wants to thank Heng Guo, Tom Hayes, Eric Vigoda, and Chaodong Zheng for helpful discussions.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Alon, N., Babai, L., Itai, A.: A fast and simple randomized parallel algorithm for the maximal independent set problem. *J. Algorithms* **7**(4), 567–583 (1986)
- Awerbuch, B., Luby, M., Goldberg, A.V., Plotkin, S.A.: Network decomposition and locality in distributed computation. In: Proceedings of the 30th IEEE Annual Symposium on Foundations of Computer Science (FOCS), pp. 364–369 (1989)
- Barenboim, L.: Deterministic $(\Delta + 1)$ -coloring in sublinear (in Δ) time in static, dynamic, and faulty networks. *J. ACM* **63**(5), 47 (2016)
- Barenboim, L., Elkin, M.: Deterministic distributed vertex coloring in polylogarithmic time. *J. ACM* **58**(5), 23 (2011)
- Barenboim, L., Elkin, M., Pettie, S., Schneider, J.: The locality of distributed symmetry breaking. *J. ACM* **63**(3), 20 (2016)
- Bubley, R., Dyer, M.: Path coupling: a technique for proving rapid mixing in markov chains. In: Proceedings of the 38th IEEE Annual Symposium on Foundations of Computer Science (FOCS), pp. 223–231 (1997)
- Cai, J.Y., Chen, X., Lu, P.: Nonnegative weighted# CSP: an effective complexity dichotomy. *SIAM J. Comput.* **45**(6), 2177–2198 (2016)
- Cai, J.Y., Galanis, A., Goldberg, L.A., Guo, H., Jerrum, M., Štefankovič, D., Vigoda, E.: # bis-hardness for 2-spin systems on bipartite bounded degree graphs in the tree non-uniqueness region. *J. Comput. Syst. Sci.* **82**(5), 690–711 (2016)

9. Calderhead, B.: A general construction for parallelizing Metropolis–Hastings algorithms. *Proc. Natl. Acad. Sci.* **111**(49), 17408–17413 (2014)
10. Chang, Y.J., Kopelowitz, T., Pettie, S.: An exponential separation between randomized and deterministic complexity in the LOCAL model. In: *Proceedings of the 57th IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 615–624 (2016)
11. Chung, K.M., Pettie, S., Su, H.H.: Distributed algorithms for the Lovász local lemma and graph coloring. In: *Proceedings of the 33rd ACM Symposium on Principles of Distributed Computing (PODC)*, pp. 134–143 (2014)
12. Dániel, M.: Graph colouring problems and their applications in scheduling. *Period. Polytech. Electr. Eng.* **48**(1–2), 11–16 (2004)
13. Das Sarma, A., Nanongkai, D., Pandurangan, G., Tetali, P.: Distributed random walks. *J. ACM* **60**(1), 2 (2013)
14. De Sa, C., Olukotun, K., Ré, C.: Ensuring rapid mixing and low bias for asynchronous Gibbs sampling. In: *Proceedings of the 33rd International Conference on Machine Learning (ICML)*, pp. 1567–1576 (2016)
15. De Sa, C., Zhang, C., Olukotun, K., Ré, C.: Rapidly mixing Gibbs sampling for a class of factor graphs using hierarchy width. In: *Advances in Neural Information Processing Systems (NIPS)*, pp. 3097–3105 (2015)
16. Dobrushin, R.L.: Prescribing a system of random variables by conditional distributions. *Theory Probab. Appl.* **15**(3), 458–486 (1970)
17. Doshi-Velez, F., Mohamed, S., Ghahramani, Z., Knowles, D.A.: Large scale nonparametric Bayesian inference: data parallelisation in the Indian buffet process. In: *Advances in Neural Information Processing Systems (NIPS)*, pp. 1294–1302 (2009)
18. Dyer, M., Goldberg, L.A., Jerrum, M.: Dobrushin conditions and systematic scan. In: *Proceedings of the 10th International Workshop on Randomization and Computation (RANDOM)*, pp. 327–338. Springer, Berlin (2006)
19. Dyer, M., Goldberg, L.A., Jerrum, M.: Systematic scan for sampling colorings. *Ann. Appl. Probab.* **16**(1), 185–230 (2006)
20. Efthymiou, C., Hayes, T.P., Štefanković, D., Vigoda, E., Yin, Y.: Convergence of MCMC and loopy BP in the tree uniqueness region for the hard-core model. In: *Proceedings of the 57th IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 704–713 (2016)
21. Feng, W., Hayes, T.P., Yin, Y.: Distributed symmetry breaking in sampling (optimal distributed randomly coloring with fewer colors). *arXiv preprint arXiv:1802.06953* (2018)
22. Feng, W., Yin, Y.: On local distributed sampling and counting. *arXiv preprint arXiv:1802.06686* (2018)
23. Fischer, M., Ghaffari, M.: A simple parallel and distributed sampling technique: local Glauber dynamics. *arXiv preprint arXiv:1802.06676* (2018)
24. Fraigniaud, P., Heinrich, M., Kosowski, A.: Local conflict coloring. In: *Proceedings of the 57th IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 625–634 (2016)
25. Fraigniaud, P., Korman, A., Peleg, D.: Towards a complexity theory for local distributed computing. *J. ACM* **60**(5), 35 (2013)
26. Frieze, A., Vigoda, E.: A survey on the use of Markov chains to randomly sample colourings. *Oxf. Lect. Ser. Math. Appl.* **34**, 53 (2007)
27. Galanis, A., Štefanković, D., Vigoda, E.: Inapproximability for antiferromagnetic spin systems in the tree nonuniqueness region. *J. ACM* **62**(6), 50 (2015)
28. Galanis, A., Štefanković, D., Vigoda, E.: Inapproximability of the partition function for the antiferromagnetic Ising and hard-core models. *Comb. Probab. Comput.* **25**(04), 500–559 (2016)
29. Ghaffari, M.: An improved distributed algorithm for maximal independent set. In: *Proceedings of the 27th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 270–277 (2016)
30. Ghaffari, M., Kuhn, F., Maus, Y.: On the complexity of local distributed graph problems. *arXiv preprint arXiv:1611.02663* (2016)
31. Ghaffari, M., Su, H.H.: Distributed degree splitting, edge coloring, and orientations. In: *Proceedings of the 28th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 2505–2523 (2017)
32. Gonzalez, J.E., Low, Y., Gretton, A., Guestrin, C.: Parallel Gibbs sampling: From colored fields to thin junction trees. In: *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS)*, vol. 15, pp. 324–332 (2011)
33. Guo, H., Jerrum, M., Liu, J.: Uniform sampling through the Lovász local lemma. In: *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pp. 342–355 (2017)
34. Harris, D.G., Schneider, J., Su, H.H.: Distributed $(\Delta + 1)$ -coloring in sublogarithmic rounds. In: *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pp. 465–478 (2016)
35. Hayes, T.P.: A simple condition implying rapid mixing of single-site dynamics on spin systems. In: *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 39–46 (2006)
36. Holroyd, A.E., Schramm, O., Wilson, D.B.: Finitary coloring. *arXiv preprint arXiv:1412.2725* (2014)
37. Jerrum, M.: A very simple algorithm for estimating the number of k -colorings of a low-degree graph. *Random Struct. Algorithms* **7**(2), 157–165 (1995)
38. Johnson, M.J., Saunderson, J., Willsky, A.S.: Analyzing Hogwild parallel Gaussian Gibbs sampling. In: *Advances in Neural Information Processing Systems (NIPS)*, pp. 2715–2723 (2013)
39. Kuhn, F., Moscibroda, T., Wattenhofer, R.: What cannot be computed locally! In: *Proceedings of the 23th Annual ACM Symposium on Principles of Distributed Computing (PODC)*, pp. 300–309 (2004)
40. Kuhn, F., Moscibroda, T., Wattenhofer, R.: The price of being near-sighted. In: *Proceedings of the 17th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 980–989. Society for Industrial and Applied Mathematics (2006)
41. Kuhn, F., Moscibroda, T., Wattenhofer, R.: Local computation: lower and upper bounds. *J. ACM* **63**(2), 17 (2016)
42. Kuhn, F., Wattenhofer, R.: On the complexity of distributed graph coloring. In: *Proceedings of the 25th Annual ACM Symposium on Principles of Distributed Computing (PODC)*, pp. 7–15 (2006)
43. Levin, D.A., Peres, Y., Wilmer, E.L.: *Markov Chains and Mixing Times*. American Mathematical Soc., Providence (2009)
44. Linial, N.: Locality in distributed graph algorithms. *SIAM J. Comput.* **21**(1), 193–201 (1992)
45. Lu, P., Yin, Y.: Improved FPTAS for multi-spin systems. In: *Proceedings of the 17th International Workshop on Randomization and Computation (RANDOM)*, pp. 639–654 (2013)
46. Luby, M.: A simple parallel algorithm for the maximal independent set problem. *SIAM J. Comput.* **15**(4), 1036–1053 (1986)
47. Mezard, M., Montanari, A.: *Information, Physics, and Computation*. Oxford University Press, Oxford (2009)
48. Moser, R.A., Tardos, G.: A constructive proof of the general Lovász local lemma. *J. ACM* **57**(2), 11 (2010)
49. Naor, M., Stockmeyer, L.: What can be computed locally? *SIAM J. Comput.* **24**(6), 1259–1277 (1995)
50. Newman, D., Asuncion, A., Smyth, P., Welling, M.: Distributed inference for latent Dirichlet allocation. In: *Proceedings of the 20th International Conference on Neural Information Processing Systems (NIPS)*, pp. 1081–1088 (2007)
51. Niu, F., Recht, B., Ré, C., Wright, S.J.: Hogwild: a lock-free approach to parallelizing stochastic gradient descent. In: *Advances in Neural Information Processing Systems (NIPS)*, pp. 693–701 (2011)

52. Peleg, D.: *Distributed Computing: A Locality-sensitive Approach*. SIAM, Philadelphia (2000)
53. Salas, J., Sokal, A.D.: Absence of phase transition for antiferromagnetic Potts models via the Dobrushin uniqueness theorem. *J. Stat. Phys.* **86**(3), 551–579 (1997)
54. Sarma, A.D., Holzer, S., Kor, L., Korman, A., Nanongkai, D., Pandurangan, G., Peleg, D., Wattenhofer, R.: Distributed verification and hardness of distributed approximation. *SIAM J. Comput.* **41**(5), 1235–1265 (2012)
55. Sly, A.: Computational transition at the uniqueness threshold. In: *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 287–296 (2010)
56. Sly, A., Sun, N.: Counting in two-spin models on d -regular graphs. *Ann. Probab.* **42**(6), 2383–2416 (2014)
57. Smyth, P., Welling, M., Asuncion, A.U.: Asynchronous distributed learning of topic models. In: *Advances in Neural Information Processing Systems (NIPS)*, pp. 81–88 (2009)
58. Swendsen, R.H., Wang, J.S.: Replica Monte Carlo simulation of spin-glasses. *Phys. Rev. Lett.* **57**(21), 2607 (1986)
59. Vigoda, E.: Improved bounds for sampling colorings. *J. Math. Phys.* **41**(3), 1555–1569 (2000)
60. Weitz, D.: Counting independent sets up to the tree threshold. In: *Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 140–149 (2006)
61. Xu, M., Lakshminarayanan, B., Teh, Y.W., Zhu, J., Zhang, B.: Distributed bayesian posterior sampling via moment sharing. In: *Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS)*, pp. 3356–3364 (2014)
62. Yan, F., Xu, N., Qi, Y.: Parallel inference for latent Dirichlet allocation on graphics processing units. In: *Advances in Neural Information Processing Systems (NIPS)*, pp. 2134–2142 (2009)
63. Yang, Y., Chen, J., Zhu, J.: Distributing the stochastic gradient sampler for large-scale LDA. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 1975–1984 (2016)