Combinatorics, Probability and Computing (2020), **29**, pp. 555–586
doi:10.1017/S0963548320000255

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Deterministic counting of graph colourings using sequences of subgraphs

Charilaos Efthymiou†

Department of Computer Science, University of Warwick, Coventry CV4 7AL, UK
Email: Charilaos.Efthymiou@warwick.ac.uk

(Received 29 September 2009; revised 14 April 2020; accepted 3 November 2012; first published online 22 June 2020)

Abstract

In this paper we propose a polynomial-time *deterministic* algorithm for approximately counting the k-colourings of the random graph $G(n, d/n)$, for constant $d > 0$. In particular, our algorithm computes in polynomial time a $(1 \pm n^{-\Omega(1)})$-approximation of the so-called 'free energy' of the k-colourings of $G(n, d/n)$, for $k \geqslant (1 + \varepsilon)d$ with probability $1 - o(1)$ over the graph instances.

Our algorithm uses *spatial correlation decay* to compute numerically estimates of marginals of the Gibbs distribution. Spatial correlation decay has been used in different counting schemes for deterministic counting. So far algorithms have exploited a certain kind of set-to-point correlation decay, e.g. the so-called *Gibbs uniqueness*. Here we deviate from this setting and exploit a point-to-point correlation decay. The spatial mixing requirement is that for a pair of vertices the correlation between their corresponding configurations becomes weaker with their distance.

Furthermore, our approach generalizes in that it allows us to compute the Gibbs marginals for small sets of nearby vertices. Also, we establish a connection between the fluctuations of the number of colourings of $G(n, d/n)$ and the fluctuations of the number of short cycles and edges in the graph.

2020 MSC Codes: 68R99, 68W25, 68W20, 82B44

1. Introduction

For a graph $G = (V, E)$ and an integer $k > 0$, a proper k-colouring is an assignment $\sigma : V \to [k]$, such that adjacent vertices receive different members of $[k]$, *i.e.* different 'colours'. Here we focus on the problem of *counting* the k-colourings of G. In particular, we consider the case where the underlying graph is an instance of the Erdős–Rényi random graph $G(n, p)$, where $p = d/n$ and d is 'large' but remains bounded as $n \to \infty$. Both in the general case and for $G(n, d/n)$, it is considered unlikely that the counting problem can be solved exactly in polynomial time. The focus is on *approximation* algorithms.

Closely related to the problem of counting colourings is the notion of the *Gibbs distribution*, *i.e.* the uniform distribution over the k-colourings of the underlying graph. Counting problems, as in the case of colouring, can be reduced very naturally to computing marginal probabilities of the Gibbs distribution [13]. In that respect, there are two basic approaches to estimating marginals of the Gibbs distribution: one is the randomized approach; the other is the deterministic approach.

†Research supported by Centre of Discrete Mathematics and Applications (DIMAP), University of Warwick, EPSRC award, and DFG grant EF 103/1-1.

© The Author(s), 2020. Published by Cambridge University Press.

https://doi.org/10.1017/S0963548320000255 Published online by Cambridge University Press

In the randomized approach, the Gibbs marginals are estimated by using a *sampling* algorithm. The most powerful method for sampling is the Markov chain Monte Carlo (MCMC) algorithm. There the main technical challenge is to establish that the underlying Markov chain mixes in polynomial time (see *e.g.* [11, 12]). For the counting algorithms which rely on MCMC sampling, the approximation guarantees are probabilistic; see [23] for MCMC sampling colouring results.

In the deterministic approach, the algorithms compute estimates of the Gibbs marginals *numerically*. Usually the deterministic algorithms exploit certain *spatial correlation decay* properties of the Gibbs distribution to reduce the load of the numerical calculations. Spatial correlation decay, or *spatial mixing*, expresses the decay of correlation between configurations of different sets of vertices with the distance between the sets. Intuitively, when we have spatial mixing, the marginals that the algorithm needs to compute are 'local quantities', *i.e.* local with respect to the graph structure. The algorithmic challenge is to exploit this locality of marginals in order to compute good approximations of them in polynomial time. There are many cases in which correlation decay is known, but there is nevertheless no known deterministic algorithm that can do the counting. Colouring seems to be one of the most challenging settings.

The connection between spatial mixing and deterministic computation of the Gibbs marginals goes far beyond the case of graph colourings; examples include the Ising model, the Potts model, the hard-core model (weighted independent sets) and many others. Even though the different models have their own special features when it comes to counting, it seems that we can deduce some basic principles for the corresponding algorithms.

Perhaps the most basic property of counting algorithms is that, not surprisingly, stronger approximation guarantees require stronger spatial mixing properties. Note that in the literature there are different notions of spatial mixing, typically of different strength. For example, the counting scheme introduced in [25], for the hard-core model on a general graph G, turns out to have the best approximation guarantees. However, this algorithm, as well as the follow-up algorithms (*e.g.* [19, 21, 22]) require particularly strong spatial mixing conditions, that is, they require what we call the *strong spatial mixing* condition. On the other hand, the weaker, in terms of approximation guarantees, algorithms in [3, 17] require a condition called *weak spatial mixing* and locally tree-like structure[1] for the underlying graphs. Strong and weak spatial mixing are fundamental concepts of spatial correlation decay. Establishing the optimal bounds where these notions hold remains open for a lot of models.

For the case where the underlying graph is $G(n, d/n)$, it is well known that rather weak spatial mixing conditions hold, *i.e.* weaker than the weak spatial mixing condition. In that respect, the question is whether we can exploit them for counting. The investigation of the interplay between spatial mixing of the Gibbs distribution and counting, for $G(n, d/n)$, becomes even more exciting in light of the mathematically non-rigorous *cavity method* [14, 15, 16]. The cavity method was introduced by statistical physicists to analyse phase transition phenomena for *random Constraint Satisfaction Problems* (r-CSP), which includes colourings of $G(n, d/n)$. Physicists involve the Gibbs distribution in their predictions, and in particular its spatial mixing properties, and this creates a connection between the cavity method and counting. Many predictions about colourings of $G(n, d/n)$ made by the cavity method have been verified rigorously; see *e.g.* [1, 5, 6, 7]. From the intuition we get from the cavity method, it is widely believed that efficient counting, with somehow reasonable approximation guarantees, is possible for $k > d/\log d$. However, it is still open how this could be done. For such a small value of k the spatial mixing conditions we have are very weak, that is, we are in the so-called 'non-reconstruction' phase.

Here we present a deterministic algorithm for computing the normalized logarithm[2] of the number of k-colourings of $G(n, d/n)$. In statistical physics terminology this quantity is known as the 'free energy'. The algorithm we propose provides an approximation up to a polynomially

[1] Strictly speaking, locally tree-like structure means the graph is of large girth.
[2] Unless otherwise specified, all the logarithms we consider are base e.

https://doi.org/10.1017/S0963548320000255 Published online by Cambridge University Press

small error, and uses the standard reduction to estimating marginals of the Gibbs distribution and spatial mixing.

What makes our approach different, compared to the previous ones, is the new way of dealing with spatial mixing. So far the algorithms have exploited a certain kind of set-to-point correlation decay; that is, under the Gibbs distribution, the correlation between the configuration of a vertex w and that of the vertices at distance r from w becomes weaker and weaker as r grows. Here we deviate from the set-to-point correlation decay and exploit a point-to-point correlation decay. The spatial mixing requirement is that for a pair of vertices the correlation between their corresponding configurations becomes weaker with their distance. The correlation decay conditions we require for our algorithm are weaker than those in the related algorithms in [3, 17]. Showing that the effect of the configuration of a single vertex on another set of vertices diminishes with the distance is equivalent to establishing non-reconstruction. This leaves open the possibility of using this approach to count k-colourings under a weaker assumption on k; of particular interest would be to count when k is smaller than d.

Our analysis of the algorithm has some very interesting corollaries too. Perhaps the most interesting one is that using the tools we developed for the analysis, we are able to trace very accurately the number of k-colourings of $G(n, d/n)$ and the source of its *fluctuations*. More specifically, we show a very specific correspondence between the fluctuations of the number of k-colourings of $G(n, d/n)$ and the fluctuations of the number of edges and the structure of short cycles in the graph by means of Gibbs correlation decay. Our result is reminiscent of the result in [4].

Subsequent work. Since the first submission of this paper, there have been some interesting subsequent results related to the randomized aspect of the problem. That is, they are mostly focusing on the related problem of generating approximately random samples of k-colourings of $G(n, d/n)$.

The author of this work, in the subsequent paper [8], uses some of the ideas that appear here to give an approximation algorithm (weak sampler) for random colourings of $G(n, d/n)$ for $k > d$. Furthermore, Yin and Zhang in [26] give a *fully polynomial-time approximate uniform sampler* (FPAUS) for generating samples of k-colourings of $G(n, d/n)$ for $k > 3d$. Finally, in a recent work with Hayes, Štefankovič and Vigoda [9], we gave an MCMC algorithm for generating random k-colourings of $G(n, d/n)$ for $k > \alpha d$, where $\alpha \approx 1.7632\ldots$ is the solution to the equation $\alpha = e^{1/\alpha}$.

Notation. For a graph $G = (V, E)$, a vertex $v \in V$ and an integer $r \geqslant 0$, we let $\texttt{Ball}_r(v)$ denote the induced subgraph of G which includes v and all the vertices which are *within distance* r from v. Also, we let $\texttt{Sphere}_r(v)$, the set of vertices which are at *distance exactly* r from v. For a vertex set $\Lambda \subseteq V$, we consider $\texttt{Sphere}_r(\Lambda)$ and $\texttt{Ball}_r(\Lambda)$ in the natural way.

We use the small letters of the Greek alphabet to indicate colourings of G, *e.g.* σ, τ, η, *etc.* For $\Lambda \subseteq V$ we use $\sigma(\Lambda)$ or σ_Λ to indicate the colour assignment of the vertices in Λ under the colouring σ.

2. Preliminaries for counting and results

For a graph $G = (V, E)$ and an integer $k > 0$, let $Z(G, k)$ be the number of (proper) k-colourings of G. In the statistical physics literature $Z(G, k)$ is also known as the *partition function*. Our focus is on computing efficiently an approximation for the log-partition function, *i.e.* $\log Z(G, k)$.

Definition 2.1. $\mathcal{Z}$ is defined to be an ε-approximation of the log-partition function $\log Z(G, k)$ if

$$(1 - \varepsilon)n^{-1} \log Z(G, k) \leqslant \mathcal{Z} \leqslant (1 + \varepsilon)n^{-1} \log Z(G, k).$$

Closely related to the partition function of the k-colourings of G is the corresponding Gibbs distribution. We let the Gibbs distribution $\mu = \mu_{G,k}$ be defined as the uniform distribution

https://doi.org/10.1017/S0963548320000255 Published online by Cambridge University Press

over the proper k-colourings of G. That is, every proper colouring σ is the assigned probability measure

$$\mu(\sigma) = \frac{1}{Z(G,k)} \quad \text{and} \quad Z(G,k) = \sum_{\sigma \in [k]^V} \mathbf{1}\{\sigma \text{ is a proper colouring of } G\}.$$

We always assume that k is sufficiently large such that the underlying graph G is k-colourable.

In the following sections we describe a *deterministic* approximation algorithm for the log-partition function when the input graph is an instance of $G(n, d/n)$. In particular we use the following standard reduction from counting to estimating the Gibbs marginals. We consider a sequence $G_0, \ldots, G_r$ of subgraphs of the input graph, where G_r is the input graph $G(n, d/n)$ and G_0 is the empty graph. There is an edge e_i in G_{i+1} such that removing it we get G_i, that is, G_{i+1} has an extra edge compared to G_i. It is elementary to show that

$$Z(G(n, d/n), k) = k^n \prod_{i=0}^{r-1} \frac{Z(G_{i+1}, k)}{Z(G_i, k)}.$$

Also, it is not hard to see that the ratio $Z(G_{i+1}, k)/Z(G_i, k)$ can be expressed in terms of the Gibbs distribution of the colourings in G_i. Letting μ_i be the aforementioned Gibbs distribution and letting $e_i = \{u, w\}$, it holds that

$$\mu_i(\mathcal{E}_i) = \frac{Z(G_{i+1}, k)}{Z(G_i, k)},$$

where $\mathcal{E}_i$ denotes the event that the two vertices u, w, the ends of the edge e_i, take on different colours.

Our aim is to show an algorithm which computes estimates of the marginals $\mu_i(\mathcal{E}_i)$ in time which is polynomial in n, when the input is a typical instance of $G(n, d/n)$. The actual estimates are within some error which tends to zero with n. More specifically, the algorithm we are going to describe has the following properties.

Theorem 2.1. *For any $\varepsilon > 0$ there exist $d_0 > 0$ such that the following is true. For any $d > d_0$ and any integer $k \geqslant (1+\varepsilon)d$, with probability at least $1 - 3n^{-\log(1+\varepsilon/4)/(52 \log d)}$ over the instances $G(n, d/n)$ there is a* deterministic *algorithm that computes an $(n^{-\log(1+\varepsilon/4)/(50 \log d)})$-approximation of $\log Z(G(n, d/n), k)$ in time $O(k^4 n^2)$.*

It is worth mentioning that for non-typical instances of $G(n, d/n)$ the algorithm in Theorem 2.1 is still a polynomial-time one, but it does not detect that the approximation guarantees are violated. See the end of Section 6 for further discussion.

The algorithm exploits correlation decay phenomena to compute estimates of the Gibbs marginals $\mu_i(\mathcal{E}_i)$, discussed above. Our main contribution, which underlies the counting algorithm, is to show how to compute good approximations of marginals of the Gibbs distribution of k-colourings of typical instances of $G(n, d/n)$, for $k > d$. Using the ideas we develop for the algorithm above, we propose a *deterministic* approximation algorithm which computes estimates of the Gibbs marginals of (relatively) small subsets of nearby vertices in $G(n, d/n)$.

Before stating our result about computing marginals, we need to introduce a notion of distance between distributions. For the distributions ν_a, ν_b on $[k]^V$, we let $\|\nu_a - \nu_b\|$ denote their *total variation distance*, that is,

$$\|\nu_a - \nu_b\| = \max_{\Omega' \subseteq [k]^V} |\nu_a(\Omega') - \nu_b(\Omega')|. \tag{2.1}$$

For $\Lambda \subseteq V$ let $\|\nu_a - \nu_b\|_\Lambda$ denote the total variation distance between the marginals of ν_a and ν_b on the vertex set Λ.

https://doi.org/10.1017/S0963548320000255 Published online by Cambridge University Press

Theorem 2.2. *For any $\varepsilon > 0$ there exists $d_0 > 0$ such that, for any $d > d_0$ and any integer $k \geqslant (1+\varepsilon)d$, with probability at least $1 - n^{-\log(1+\varepsilon/4)/(20\log d)}$ over the instances of $G(n, d/n)$, the following is true.*

Let Λ be any vertex set in $G(n, d/n)$ such that $|\Lambda| = o(\log n)$ and with maximum distance between any two vertices in Λ being $o(\log n)$. There is a deterministic *algorithm that, in $O(|\Lambda| k^3 n)$ steps, computes ν, an approximation of the Gibbs marginal on Λ, such that*

$$\|\mu_\Lambda - \nu\| \leqslant n^{-\log(1+\varepsilon/4)/(20\log d)},$$

where μ_Λ is the Gibbs marginal on Λ.

The proof of Theorem 2.2 appears in Section 7.

As before, for non-typical instances of $G(n, d/n)$ the algorithm in Theorem 2.2 is still a polynomial-time one, but it cannot detect that the approximation guarantees are violated. See Section 7 for further details.

A corollary from the analysis of the above results is that the number of k-colourings of $G(n, d/n)$ is concentrated in a very precise sense. We get a direct relation between the fluctuation of the number of k-colourings of $G(n, d/n)$ and the fluctuations of the number of edges and the short-cycle structure of $G(n, d/n)$. This is reminiscent of the so-called 'small subgraph conditioning' technique [10, 20] and particularly [4].

In order to state Theorem 2.3 we introduce the following notation: for each integer $k > 0$, define

$$\mathcal{F}_{k,\ell}(G(n, d/n)) = \sum_{j=3}^{\ell-1} K_j \log\left(1 + \frac{(-1)^j}{(k-1)^{j-1}}\right),$$

where K_j is the number of cycles of length j in $G(n, d/n)$.

Theorem 2.3. *For every $\varepsilon > 0$, there exists $d_0 > 0$ such that, for $d > d_0$ and each integer $k \geqslant (1+\varepsilon)d$, the following is true. For*

$$(2\varepsilon^{-2}\log k) \leqslant \ell \leqslant \frac{1}{12}\frac{\log n}{\log d},$$

with probability at least $1 - 30k^2 d\varepsilon^{-1}(1+\varepsilon/4)^{-\ell/2}$ over the instances $G(n, d/n)$, we have

$$|\log(Z(G(n, d/n), k) - n\log k - m\log(1-k^{-1}) - \mathcal{F}_{k,\ell}(G(n, d/n))| \leqslant (1+\varepsilon/4)^{-\ell/2},$$

where m is the number of edges in $G(n, d/n)$.

The proof of Theorem 2.3 appears in Section 8.

We remark that the above result compares $\log Z(G(n, d/n), k)$ with quantities that are inherently random, *i.e.* the number of edges m and the quantity $\mathcal{F}_{k,\ell}(G(n, d/n))$. Note that the latter quantity depends on the number of the cycles of length less than ℓ in $G(n, d/n)$. Also, note that the concentration result is with respect to the *non-normalized* log partition function.

The dependence of the log-partition function on the number of short cycles of $G(n, d/n)$ is somehow natural. For $k \geqslant (1+\varepsilon)d$, Theorem 2.3 is a stronger and more explicit version of the result in [4], which is subsequent to this paper. The result in [4] uses the small subgraph conditioning technique to study the concentration of the log-partition function of $G(n, m)$ with expected degree d, *i.e.* the random graph on n vertices and $m = dn/2$ edges. The concentration bound in [4] is alas somewhat weaker, but applies over a larger range of k, *i.e.* k greater than $\approx d/(2\log d)$.

Since, typically, the number of edges in $G(n, d/n)$ is in the range $dn/2 \pm n^{2/3}$, setting $\ell = \log n/(12\log d)$, we get the following corollary from Theorem 2.3.

https://doi.org/10.1017/S0963548320000255 Published online by Cambridge University Press

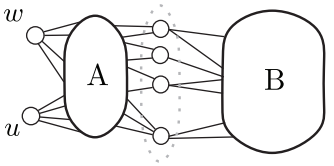


Figure 1. $\Lambda = \{v, w\}$. S is the vertices within the dotted line.

Corollary 2.1. *For every $\varepsilon > 0$, there exists $d_0 > 0$ such that, for any $d > d_0$ and any integer $k \geqslant (1+\varepsilon)d$, the following is true. With probability at least $1 - n^{-\log(1+\varepsilon/4)/(25\log d)}$ over the graph instances of $G(n, d/n)$, we have*

$$|n^{-1} \log Z(G(n, d/n), k) - (\log k + (d/2)\ \log(1 - k^{-1}))| \leqslant n^{-1/4}.$$

2.1 Structure of the paper

The rest of the paper is organized as follows. In Section 3 we provide a high-level description of how we exploit correlation decay for counting. In Section 4 we give a general description of our counting algorithm and relate its accuracy with certain quantities which express spatial correlation decay. Then, in Section 5, we provide a technique for bounding spatial correlation decay. In Section 6 we discuss the technical details for applying the counting algorithm on $G(n, d/n)$ and we prove Theorem 2.1. Finally, in Section 7 we prove Theorem 2.2 and in Section 8 we prove Theorem 2.3.

3. Computing Gibbs marginals: preliminaries

In what follows we discuss our approach to computing approximations of the Gibbs marginals. We describe how we exploit spatial mixing to compute estimates of marginals of the Gibbs distribution efficiently for graphs which are locally tree-like. In particular, we assume that we are dealing *exclusively* with locally tree-like graphs, that is, there exists some integer r such that the radius r neighbourhood of each vertex in the graph is a tree.

Let $k > 0$ be an integer and let $G = (V, E)$ be a k-colourable graph which is locally tree-like. Also, let μ be the Gibbs distribution over the k-colourings of G, *i.e.* the uniform distribution over the k-colourings.

In what follows we adopt the convention that for some $S, J \subset V$ and $\sigma \in [k]^S$, we let

$$\mu(\sigma) = \sum_{\tau \in [k]^V} \mathbf{1}\{\sigma(u) = \tau(u) \text{ for all } u \in \Lambda\}\mu(\tau). \tag{3.1}$$

Furthermore, for $\tau \in [k]^J$ and $\sigma \in [k]^S$ the conditional distribution $\mu(\sigma \mid \tau)$ has the natural interpretation.

One possible approach to computing an approximation of μ for locally tree graphs, like the one we consider here, was suggested in [3]. Roughly speaking, the idea is as follows. Assume that we want to compute the Gibbs marginal on $\Lambda = \{u, w\}$, for two vertices u, w. This marginal can be expressed as a convex combination of boundary conditions on $S = \mathtt{Sphere}_r(\Lambda)$. That is, for any $\sigma \in [k]^\Lambda$ we have

$$\mu(\sigma) = \sum_{\tau \in [k]^S} \mu(\sigma \mid \tau)\mu(\tau), \tag{3.2}$$

(see *e.g.* Figure 1). Clearly, fixing the colouring of S, the quantity $\mu(\sigma \mid \tau)$ depends only on the set of vertices which are within distance r from Λ. In Figure 1, given the colouring of the vertices in

https://doi.org/10.1017/S0963548320000255 Published online by Cambridge University Press

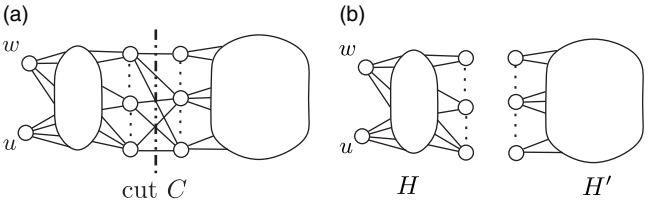


Figure 2. (a) $\Lambda = \{v, w\}$. (b) Simplification.

S, the configurations in parts A and B are independent from each other. The task of the algorithm is to compute $\mu(\sigma \mid \tau)$, $\mu(\tau)$ for various $\tau \in [k]^S$.

As far as $\mu(\tau)$ is concerned, in general we do not expect its computation to be much easier than actually computing $\mu(\sigma)$, *i.e.* our initial problem. Bandyopadhyay and Gamarnik [3] noticed that the problem of estimating coefficients $\mu(\tau)$ somehow 'degenerates' when there is a certain kind of asymptotic independence between the configurations at Λ and S. That is, we have $\mu(\sigma \mid \tau) \approx \mu(\sigma \mid \tau')$, where τ, τ' are any two colourings of S. If the aforementioned independence holds for μ, then trivially we have $\mu(\sigma) \approx \mu(\sigma \mid \tau)$. Then the problem just reduces to computing $\mu(\sigma \mid \tau)$.

For locally tree-like graphs, choosing r appropriately, the subgraph of G which $\mu(\sigma \mid \tau)$ depends on is a tree. For example, in Figure 1 we would choose r so that part A is tree-like. Then one can compute $\mu(\sigma \mid \tau)$ by only considering the tree-like neighbourhood around Λ. Such a computation can be carried out efficiently by using standard methods, *e.g. dynamic programming* [24].

Our approach here is in a similar spirit. As before, consider G and the set $\Lambda = \{v, w\}$. We let H be the induced subgraph of G that contains Λ and all its neighbours within graph distance r. For the new graph H, we let $\bar{\mu}$ be the Gibbs distribution. Our approach amounts to approximating $\mu(\sigma)$ with $\bar{\mu}(\sigma)$, for every $\sigma \in [k]^\Lambda$. Perhaps Figure 2 illustrates our approach better. Consider the graph in Figure 2(a), where $\Lambda = \{v, w\}$, and we want to compute the probability of $\sigma \in [k]^\Lambda$. As Figure 2(b) illustrates, H emerges from the deletion of all the edges in the cut C.

The idea that underlies our approach is on one hand to choose r in a such a way that H is sufficiently simply structured and allows an efficient computation for $\bar{\mu}(\sigma)$, while on the other hand $\bar{\mu}(\sigma)$ approximates $\mu(\sigma)$ well.

A priori the distributions μ and $\bar{\mu}$ are different since the underlying graphs are different. The key question is how well $\bar{\mu}(\sigma)$ approximates $\mu(\sigma)$. The distribution $\bar{\mu}(\,\cdot\,)$ can be perceived as being induced by the deletion of the edges in G that connect $\texttt{Ball}_r(\Lambda)$ with the rest of the graph. For $\bar{\mu}(\sigma)$ to be close to $\mu(\sigma)$, we require that the deletion of these edges does not have a large effect on the marginals on Λ. The comparison of the two marginals is by means of examining the influence of the Gibbs marginals on Λ we get by deleting each edge in the cut C. From a technical perspective, edge deletions allow more flexibility for manipulations, since we always consider the free Gibbs measures.

4. Counting schema

For the sake of definiteness, we present the counting schema by assuming that we are given a fixed graph $G = (V, E)$ and an integer k such that $Z(G, k) > 0$. A concept we use is that of the *sequence of subgraphs*.

Definition 4.1 (sequence of subgraphs)**.** For the graph $G = (V, E)$, let $\mathcal{G}(G) = \{G_i = (V, E_i)\}_{i=0}^r$ denote a sequence of subgraphs of G which has the following properties.

- G_0 is a spanning subgraph of G.
- $E_i \subset E_{i+1}$ for $0 \leqslant i < r$ and $E_r = E$.
- The term G_{i+1} compared to G_i has the additional edge $e_i = \{w_i, u_i\}$.

https://doi.org/10.1017/S0963548320000255 Published online by Cambridge University Press

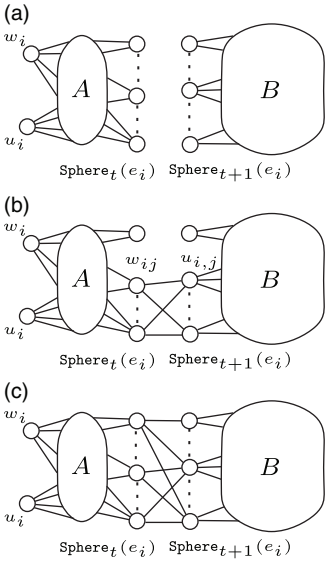


Figure 3. (a) Graph G_{i0}. (b) Graph $G_{i,j+1}$. (c) Graph G_{i,r_i}.

When we refer to $\mathcal{G}(G)$ we specify the graph G_0 and the rule which gives the terms $G_1, \ldots, G_r$. For this section we assume that G_0 is empty. For the rule which decides the order in which we add edges in the sequence, the reader may assume one, arbitrarily.

For each $G_i \in \mathcal{G}(G)$, let μ_i be the corresponding Gibbs distribution. Furthermore, let $\mathcal{E}_i$ be the event that vertices w_i, u_i take on a different colours. The following, folklore result relates the Gibbs marginals μ_i with $Z(G, k)$. Its proof is standard and for this reason we omit it.

Lemma 4.1. *For any integer $k > 0$ and any k-colourable $G = (V, E)$, consider a sequence of subgraphs $\mathcal{G}(G)$ where G_0 is empty. It holds that*

$$Z(G, k) = k^n \prod_{i=0}^{|E|-1} \mu_i(\mathcal{E}_i).$$

Our focus is on the computation of $\mu_i(\mathcal{E}_i)$ for $i = 0, \ldots, r-1$. For each $G_i \in \mathcal{G}(G)$, consider a new sequence of subgraphs $\mathcal{G}(G_i) = G_{i,0}, \ldots, G_{i,r_i}$ defined as follows. The term G_{i,r_i} is the graph G_i, while $G_{i,0}$ is derived from G_i by removing all the edges at distance t from e_i, *i.e.* the edges between the vertex sets $\mathtt{Sphere}_t(e_i)$ and $\mathtt{Sphere}_{t+1}(e_i)$. The distance t at which we delete the edges is chosen appropriately. As before, the reader may choose an arbitrary order of the edges to be deleted in $\mathcal{G}(G_i)$.

Figure 3(a–c) illustrates some members of $\mathcal{G}(G_i)$. Figure 3(a) shows the first term of the sequence. Figure 3(b) shows the graph $G_{i,j+1}$, *i.e.* the edge $e_{i,j} = \{w_{i,j}, u_{u,i}\}$ has just been inserted. In Figure 3(c) we have the final term of $\mathcal{G}(G_i)$, the graph G_{i,r_i}.

Let $\mu_{i,j}$ be the Gibbs distribution specified by the graph $G_{i,j}$. The scheme amounts to approximating $\mu_i(\mathcal{E}_i)$ by using $\mu_{i,0}(\mathcal{E}_i)$. This approximation of Gibbs marginals gives rise to approximating $Z(G, k)$ by the following quantity:

$$\mathcal{Z}(G, k) = k^n \prod_{i=0}^{|E|-1} \mu_{i,0}(\mathcal{E}_i). \tag{4.1}$$

Note that each $\mu_{i,0}()$ depends only on the configuration of vertices within distance t from the edge e_i. We choose t appropriately such that the computation of $\mu_{i,0}(\mathcal{E}_i)$ can be carried out in

https://doi.org/10.1017/S0963548320000255 Published online by Cambridge University Press

polynomial time. For the cases we consider here, we need to compute the marginal $\mu_{i,0}(\mathcal{E}_i)$ on graphs which are either trees or unicyclic. For the trees we can use dynamic programming to compute the corresponding marginals. For the cases where the graph is unicyclic, our approach is not very different from that of the tree, that is, there is an easy way of circumventing the issue of the cycle; see Section 6 for further technical details.

4.1 Accuracy of counting

We now focus on the accuracy of the aforementioned approximation scheme. In particular, our aim is to express approximation guarantees of our counting schema in terms of spatial correlation decay of the Gibbs distributions $\mu_{i,j}$. In order to get there, we first obtain the following result.

Proposition 4.2. *For a graph $G=(V,E)$ consider some arbitrary $\mathcal{G}(G)$ such that G_0 is empty. Let $r+1$ be the number of terms in the sequence, while ν_i is an estimation of the quantity $\mu_i(\mathcal{E}_i)$, for $i=0,\dots,r-1$. Letting $\mathcal{Z}=k^n\prod_{i=0}^{r-1}\nu_i$, we have*

$$\sum_{i=0}^{r-1}\frac{\nu_i-\mu_i(\mathcal{E}_i)}{\nu_i}\leqslant\log\mathcal{Z}-\log Z(G,k)\leqslant\sum_{i=0}^{r-1}\frac{\nu_i-\mu_i(\mathcal{E}_i)}{\mu_i(\mathcal{E}_i)}.$$

Proof. From the definition of our quantities we have

$$\log\mathcal{Z}-\log Z(G,k)=\sum_{i=0}^{r-1}\log\frac{\nu_i}{\mu_i(\mathcal{E}_i)}. \tag{4.2}$$

We also have

$$\log\frac{\nu_i}{\mu_i(\mathcal{E}_i)}=\log\left(1+\frac{\nu_i-\mu_i(\mathcal{E}_i)}{\mu_i(\mathcal{E}_i)}\right)\leqslant\frac{\nu_i-\mu_i(\mathcal{E}_i)}{\mu_i(\mathcal{E}_i)}, \tag{4.3}$$

where in the last inequality we use the fact that $\log(1+x)\leqslant x$. The upper bound for $\log\mathcal{Z}-\log Z(G,k)$ follows by plugging (4.3) into (4.2). As far as the lower bound is concerned, we have

$$\log\frac{\nu_i}{\mu_i(\mathcal{E}_i)}=-\log\frac{\mu_i(\mathcal{E}_i)}{\nu_i}=-\log\left(1+\frac{\mu_i(\mathcal{E}_i)-\nu_i}{\nu_i}\right)\geqslant-\frac{\mu_i(\mathcal{E}_i)-\nu_i}{\nu_i}, \tag{4.4}$$

where in the last inequality we also used the inequality $\log(1+x)\leqslant x$. The lower bound follows by plugging the above inequality into (4.2). The proposition follows. □

We use Proposition 4.2 to bound

$$n^{-1}|\log\mathcal{Z}(G,k)-\log Z(G,k)|,$$

where $\mathcal{Z}(G,k)$ is defined in (4.1). In our setting, it turns out that both $\mu_i(\mathcal{E}_i)$ and its approximation $\mu_{i,0}(\mathcal{E}_i)$ are bounded away from zero. Our focus then is on bounding $|\mu_i(\mathcal{E}_i)-\mu_{i,0}(\mathcal{E}_i)|$, for every $i=0,\dots,r-1$. The definition of total variation distance and the triangle inequality yield the following: for any $i=0,\dots,r-1$, we have

$$|\mu_i(\mathcal{E}_i)-\mu_{i,0}(\mathcal{E}_i)|\leqslant\|\mu_i(\,\cdot\,)-\mu_{i,0}(\,\cdot\,)\|_{e_i}\leqslant\sum_{j=0}^{r_i-1}\|\mu_{i,j}(\,\cdot\,)-\mu_{i,j+1}(\,\cdot\,)\|_{e_i},$$

where with a slight abuse of notation we regard the edge $e_i=\{u_i,w_i\}$ to be the set of its incident vertices.

Note that the graphs $G_{i,j}$ and $G_{i,j+1}$, with respect to which we consider the total variation distances on the right-hand side of the inequality above, differ only on the edge $e_{i,j}$. In a sense

https://doi.org/10.1017/S0963548320000255 Published online by Cambridge University Press

the quantity $\|\mu_{i,j}(\,\cdot\,) - \mu_{i,j+1}(\,\cdot\,)\|_{e_i}$ expresses the effect of the removal of edge $e_{i,j}$ on the joint distributions of the configuration on u_i, w_i.

To bound each of the terms $\|\mu_{i,j}(\,\cdot\,) - \mu_{i,j+1}(\,\cdot\,)\|_{e_i}$ above, we use the following lemma.

Lemma 4.3. *For $0 \leqslant i \leqslant r-1$ and $0 \leqslant j \leqslant r_i - 1$, the following is true. Let*

$$\mathcal{A}_{i,j} = \max_{\sigma,\tau\in[k]^{w_{i,j}}} \|\mu_{i,j}(\cdot \mid \sigma) - \mu_{i,j}(\cdot \mid \tau)\|_{e_i\cup\{u_{i,j}\}}$$

and

$$\mathcal{B}_{i,j} = \max_{\sigma,\tau\in[k]^{w_{i,j}}} \|\mu_{i,j}(\cdot \mid \sigma) - \mu_{i,j}(\cdot \mid \tau)\|_{\{u_{i,j}\}},$$

where we recall that $\mu_{i,j}$ is the Gibbs distribution over the k-colourings of $G_{i,j}$. Finally, let

$$\mathcal{C}_{i,j} = 2 \max_{\sigma\in[k]^{e_{i,j}}} \left\{(\mu_{i,j}(\sigma(u_{i,j}) \mid \sigma(w_{i,j})))^{-2}\right\}.$$

It holds that

$$\|\mu_{i,j}(\,\cdot\,) - \mu_{i,j+1}(\,\cdot\,)\|_{e_i} \leqslant \mathcal{C}_{i,j}\left(\mathcal{A}_{i,j} + \mathcal{B}_{i,j}\right).$$

Before giving the proof of Lemma 4.3, perhaps it is useful to describe the most important quantities in the lemma, *i.e.* $\mathcal{A}_{i,j}$ and $\mathcal{B}_{i,j}$. For $\mathcal{A}_{i,j}$ we fix the configuration at $w_{i,j}$ with σ and τ, and compare the corresponding marginal on the set $S = \{w_i, u_i, u_{i,j}\}$. Intuitively, the quantity $\mathcal{A}_{i,j}$ expresses the strength of the correlation between the configuration at vertex $w_{i,j}$ and that at S. Similarly, $\mathcal{B}_{i,j}$ expresses the strength of the correlation between the configuration at vertex $w_{i,j}$ and that at $u_{i,j}$.

Proof of Lemma 4.3. For *any* coupling $\phi\colon [k]^{e_{i,j}} \times [k]^{e_{i,j}} \to [0,1]$ of the marginals of $\mu_{i,j}$ and $\mu_{i,j+1}$ on the vertices $u_{i,j}, w_{i,j}$, we have the following:

$$\begin{aligned}
\|\mu_{i,j}(\,\cdot\,) - \mu_{i,j+1}(\,\cdot\,)\|_{e_i} &= \frac{1}{2}\sum_{\eta\in[k]^{e_i}} |\mu_{i,j}(\eta) - \mu_{i,j+1}(\eta)| \\
&= \frac{1}{2}\sum_{\eta\in[k]^{e_i}} \Big|\sum_{\sigma,\tau\in[k]^{e_{ij}}} (\mu_{i,j}(\eta \mid \sigma) - \mu_{i,j+1}(\eta \mid \tau))\phi(\sigma,\tau)\Big| \\
&\leqslant \frac{1}{2}\sum_{\eta\in[k]^{e_i}} \sum_{\sigma,\tau\in[k]^{e_{ij}}} |\mu_{i,j}(\eta \mid \sigma) - \mu_{i,j+1}(\eta \mid \tau)|\phi(\sigma,\tau) \\
&= \sum_{\sigma,\tau\in[k]^{e_{ij}}} \phi(\sigma,\tau)\,\frac{1}{2}\sum_{\eta\in[k]^{e_i}} |\mu_{i,j}(\eta \mid \sigma) - \mu_{i,j+1}(\eta \mid \tau)| \\
&= \sum_{\sigma,\tau\in[k]^{e_{ij}}} \phi(\sigma,\tau)\|\mu_{i,j}(\cdot \mid \sigma) - \mu_{i,j+1}(\cdot \mid \tau)\|_{e_i} \\
&\leqslant \max_{(\sigma,\tau)\in\mathrm{supp}(\phi)} \|\mu_{i,j}(\cdot \mid \sigma) - \mu_{i,j+1}(\cdot \mid \tau)\|_{e_i}, \qquad (4.5)
\end{aligned}$$

where $\mathrm{supp}(\phi)$ is the support of ϕ. In the first line we employ the definition of total variation distance. In the second line we employ the law of total probability. The last inequality follows from the observation that the convex combination of the total variation distances is at most the maximum among them.

https://doi.org/10.1017/S0963548320000255 Published online by Cambridge University Press

For this proof we specify the coupling ϕ to be as follows: we couple the configuration at the vertices in $e_{i,j} = \{u_{i,j}, w_{i,j}\}$ sequentially. First we specify the coupling for the configuration at $u_{i,j}$. Then, given the outcome of the coupling at $u_{i,j}$, we specify the coupling for $w_{i,j}$. The marginal of the configuration at $u_{i,j}$ for both $\mu_{i,j}$ and $\mu_{i,j+1}$ is the uniform distribution over $[k]$. In that respect, we specify that ϕ is the identity coupling for the vertex $u_{i,j}$. Furthermore, given the colour assignment of $u_{i,j}$ in the coupling, we couple the configuration at $w_{i,j}$ such that the probability that the vertex takes on different colours is minimized.

For the above coupling we have $(\sigma, \tau) \in \mathrm{supp}(\phi)$ only if $\sigma(u_{i,j}) = \tau(u_{i,j})$. On the other hand we may have $\sigma(w_{i,j}) \neq \tau(w_{i,j})$. From the above we have

$$\begin{aligned}\|\mu_{i,j}(\,\cdot\,) - \mu_{i,j+1}(\,\cdot\,)\|_{e_i} &\leqslant \max_{\sigma,\tau\,:\,\sigma(u_{i,j})=\tau(u_{i,j})} \|\mu_{i,j}(\cdot \mid \sigma) - \mu_{i,j+1}(\cdot \mid \tau)\|_{e_i} \\ &= \max_{\sigma,\tau\,:\,\sigma(u_{i,j})=\tau(u_{i,j})} \|\mu_{i,j}(\cdot \mid \sigma) - \mu_{i,j}(\cdot \mid \tau)\|_{e_i}, \end{aligned} \tag{4.6}$$

where σ, τ vary over the configuration of $u_{i,j}, w_{i,j}$. The second derivation comes from the observation that for any configuration τ at the vertices $u_{i,j}$ and $w_{i,j}$, the measures $\mu_{i,j+1}(\cdot \mid \tau)$ and $\mu_{i,j}(\cdot \mid \tau)$ are identical.

By the definition of total variation distance, there exists $\mathcal{S} \subseteq [k]^{e_i}$ such that

$$\|\mu_{i,j}(\cdot \mid \tau) - \mu_{i,j}(\cdot \mid \sigma)\|_{e_i} = |\mu_{i,j}(\mathcal{S} \mid \tau) - \mu_{i,j}(\mathcal{S} \mid \sigma)|. \tag{4.7}$$

Letting $E_{ij} = \mu_{ij}(\tau_{u_{ij}}|\tau_{w_{ij}}) - \mu_{ij}(\sigma_{u_{ij}}|\sigma_{w_{ij}})$, (4.7) implies that

$$\begin{aligned}\|\mu_{i,j}(\cdot \mid \tau) - \mu_{i,j}(\cdot \mid \sigma)\|_{e_i} &= |\mu_{i,j}(\mathcal{S} \mid \tau) - \mu_{i,j}(\mathcal{S} \mid \sigma)| \\ &\leqslant \left| \frac{\mu_{i,j}\big(\mathcal{S}, \tau_{u_{ij}}|\tau_{w_{ij}}\big)}{\mu_{i,j}\big(\tau_{u_{ij}}|\tau_{w_{ij}}\big)} - \frac{\mu_{i,j}\big(\mathcal{S}, \sigma_{u_{ij}}|\sigma_{w_{ij}}\big)}{\mu_{i,j}\big(\sigma_{u_{ij}}|\sigma_{w_{ij}}\big)} \right| \\ &\leqslant \left| \frac{\mu_{i,j}\big(\mathcal{S}, \tau_{u_{ij}}|\tau_{w_{ij}}\big)}{\mu_{i,j}\big(\sigma_{u_{ij}}|\sigma_{w_{ij}}\big) + E_{ij}} - \frac{\mu_{i,j}\big(\mathcal{S}, \sigma_{u_{ij}}|\sigma_{w_{ij}}\big)}{\mu_{i,j}\big(\sigma_{u_{ij}}|\sigma_{w_{ij}}\big)} \right| \\ &\leqslant \frac{|\mu_{i,j}\big(\mathcal{S}, \tau_{u_{ij}}|\tau_{w_{ij}}\big) - \mu_{i,j}\big(\mathcal{S}, \sigma_{u_{ij}}|\sigma_{w_{ij}}\big)|}{\mu_{i,j}\big(\sigma_{u_{ij}}|\sigma_{w_{ij}}\big)} + \frac{|E_{i,j}|}{\mu_{i,j}\big(\tau_{u_{ij}}|\tau_{w_{ij}}\big)\mu_{i,j}\big(\sigma_{u_{ij}}|\sigma_{w_{ij}}\big)} \\ &\leqslant \frac{|\mu_{i,j}\big(\mathcal{S}, \tau_{u_{ij}}|\tau_{w_{ij}}\big) - \mu_{i,j}\big(\mathcal{S}, \sigma_{u_{ij}}|\sigma_{w_{ij}}\big)| + |E_{i,j}|}{\mu_{i,j}\big(\tau_{u_{ij}}|\tau_{w_{ij}}\big)\mu_{i,j}\big(\sigma_{u_{ij}}|\sigma_{w_{ij}}\big)}. \end{aligned}$$

The lemma follows by using the above, the assumption that $\tau(u_{i,j}) = \sigma(u_{i,j})$ and noting that the following inequalities hold:

$$\begin{aligned}|\mu_{i,j}\big(\mathcal{S}, \tau_{u_{ij}}|\tau_{w_{ij}}\big) - \mu_{i,j}\big(\mathcal{S}, \sigma_{u_{ij}}|\sigma_{w_{ij}}\big)| &\leqslant \|\mu_{i,j}\big(\,\cdot\,|\tau_{w_{ij}}\big) - \mu_{i,j}\big(\,\cdot\,|\sigma_{w_{ij}}\big)\|_{e_i \cup \{u_{i,j}\}} \leqslant \mathcal{A}_{i,j}, \\ |E_{ij}| &\leqslant 2\,\|\mu_{i,j}\big(\,\cdot\,|\tau_{w_{ij}}\big) - \mu_{i,j}\big(\,\cdot\,|\sigma_{w_{ij}}\big)\|_{u_{ij}} \leqslant 2\mathcal{B}_{i,j}, \end{aligned}$$

while

$$2(C_{i,j})^{-1} \leqslant \mu_{i,j}(\tau_{u_{ij}}|\tau_{w_{ij}})\mu_{i,j}(\sigma_{u_{ij}}|\sigma_{w_{ij}}). \qquad \square$$

5. Bounds for spatial correlation decay

In this section we present our approach for bounding the quantities that express the spatial correlation decay we introduced in Section 4.1. Given a graph $G = (V, E)$, $\Lambda \subset V$ and some vertex $w \notin \Lambda$, we want to study the 'influence' of the configuration at w on the distribution of the configuration in Λ. To be more specific, we consider the two conditional Gibbs distributions $\mu(\cdot \mid \sigma)$

https://doi.org/10.1017/S0963548320000255 Published online by Cambridge University Press

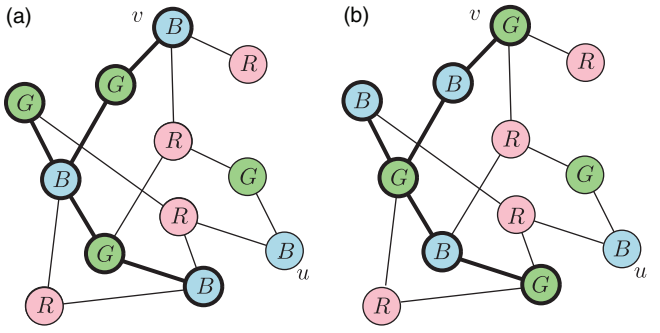


Figure 4. (a) Colouring X. (b) Colouring Y.

and $\mu(\cdot \mid \tau)$, where σ and τ are two configurations at vertex w, respectively. The aim here is to bound the total variation distance $\|\mu(\cdot \mid \sigma) - \mu(\cdot \mid \tau)\|_\Lambda$. In what follows we present and prove Theorem 5.1, which is the main contribution of this section.

Theorem 5.1. *Consider the graph $G = (V, E)$, $w \in V$, $\Lambda \subseteq V$ and an integer $k > 0$. Let μ denote the Gibbs distribution of the k-colourings of G. Also, for any two $c, q \in [k]^w$, different from each other, let $\mathcal{B}_{c,q}$ be the event that there is a path from w to Λ such that all the vertices in that path are coloured only with the colours c, q. We have*

$$\|\mu(\cdot \mid c) - \mu(\cdot \mid q)\|_\Lambda \leqslant \mu(\mathcal{B}_{c,q} \mid c).$$

Note that using Theorem 5.1 we can have a bound on the quantities $\mathcal{A}_{i,j}$ and $\mathcal{B}_{i,j}$ of Lemma 4.3, in terms of probabilities of paths of disagreement getting specific colours. For $\mathcal{A}_{i,j}$, for example, the variables $w_{i,j}$ and $\{w_i, u_i, u_{i,j}\}$ correspond to the variables w, Λ of Theorem 5.1, respectively.

Our approach to proving Theorem 5.1 is to use coupling. We consider X, Y distributed as in $\mu(\cdot \mid c)$, $\mu(\cdot \mid q)$, respectively, and we couple these two random variables. By the Coupling Lemma [2] we have

$$\|\mu(\cdot \mid c) - \mu(\cdot \mid q)\|_\Lambda \leqslant \mathbb{P}[X(\Lambda) \neq Y(\Lambda)]. \tag{5.1}$$

We define the coupling by specifying Y, given X. More specifically, given X, we let G_X denote the maximal *connected* subgraph of G which contains the vertex w and vertices from the colour classes specified by c and q in the colouring X. Then we derive Y as follows. For every vertex $u \notin G_X$ we have $Y(u) = X(u)$. For $u \in G_X$, if $X(u) = c$, then $Y(u) = q$, while if $X(u) = q$, then $Y(u) = c$. In Figure 4(a, b) we illustrate this coupling, *e.g.* $c =$'blue' and $q =$'green'.

Proposition 5.1. *The above is a valid coupling between the distributions $\mu(\cdot \mid c)$ and $\mu(\cdot \mid q)$.*

Proof. Let Ω_c and Ω_q denote the k-colourings of G that assign the vertex v colour c and q, respectively. We claim that the coupling we describe above can be viewed as a map from the set Ω_c to the set Ω_q. That is, our coupling is defined as a map $T_{c,q}$ such that $T_{c,q}\colon \Omega_c \to \Omega_q$. To prove this claim we need to show that for every $\sigma \in \Omega_c$ we have $T_{c,q}(\sigma) \in \Omega_q$.

For any $\sigma \in \Omega_c$, consider $G_\sigma = (V_\sigma, E_\sigma)$ as defined above. We need to focus on three properties of G_σ. The first is that it is bipartite (in the extreme case where $V_\sigma = \{w\}$ we consider G_σ bipartite too). The second is that G_σ is connected. The third is that G_σ is maximal. This implies the following: if $\partial V_\sigma = \{v \in V \setminus V_\sigma \mid \exists \{v, u\} \in E \text{ with } u \in V_\sigma\}$, then for all $v \in \partial V_\sigma$ it holds that $\sigma(v) \notin \{c, q\}$.

Clearly σ specifies a proper 2-colouring for the vertices of G_σ that uses only the colours c and q. In particular, let $p_1, p_2 \subseteq V_\sigma$ be the two parts of G_σ and without loss of generality assume that v belongs to p_1. Then σ assigns to all the vertices in p_1 the colour c and to all the vertices in p_2 the colour q. In those terms, the mapping $T_{c,q}$ works as follows. For every vertex $v \in V \setminus V_\sigma$ it

https://doi.org/10.1017/S0963548320000255 Published online by Cambridge University Press

holds that $T(\sigma)(v) = \sigma(v)$. For the vertices that belong to G_σ, the mapping $T_{c,q}$ swaps the colour assignments of the two parts of G_σ.

First we show that $T_{c,q}$ maps every colouring of Ω_c to Ω_q. For some $\sigma \in \Omega_c$, let $\eta = T_{c,q}(\sigma)$. It is direct that $\eta(w) = q$. It remains to show that η is a proper colouring of G. If η is a non-proper colouring, then there should be at least two adjacent vertices having the same colour assignment. Since $T_{c,q}$ recolours only the vertices in V_σ if η is a non-proper colouring, then the monochromatic pair of adjacent vertices has either both vertices in V_σ or one vertex in V_σ and the other in ∂V_σ.

It is direct that swapping the colour assignments of the two parts of G_σ leads to a proper colouring of G_σ. Thus, in η there is no monochromatic pair both of whose vertices belong to G_σ. Also, this swap of colourings cannot lead some vertex in V_σ to have the same colour assignment with some vertex in ∂V_σ. This is due to the maximality of G_σ, that is, the colouring σ cannot specify a colour assignment that uses the colours c and q for any vertex in ∂V_σ. Thus η is a proper colouring of G.

We have proved that indeed $T_{c,q}\colon \Omega_c \to \Omega_q$. For the proposition it suffices to show that the map $T_{c,q}\colon \Omega_c \to \Omega_q$ is a *bijection*. For this note that the sets Ω_c and Ω_q are equally sized. Clearly this implies that Ω_c and Ω_q admit a bijection.

In order to show that $T_{c,q}$ is indeed a bijection, we use the observation that $T_{q,c}$ is the inverse map of $T_{c,q}$. That is, it is direct that for any $\sigma \in \Omega_c$ and $\eta \in \Omega_q$ such that $\eta = T_{c,q}(\sigma)$, it holds that $\sigma = T_{q,c}(\eta)$. Then it is clear that $T_{c,q}$ is surjective, since each $\eta \in \Omega_q$ is the image of $T_{q,c}(\eta)$. Furthermore, $T_{c,q}$ is one-to-one. If we have σ_1, σ_2 such that $\eta = T_{c,q}(\sigma_1) = T_{c,q}(\sigma_2)$, then there should be that $\sigma_1 = \sigma_2$ since $\sigma_1 = T_{q,c}(\eta) = \sigma_2$. Since $T_{c,q}$ is surjective and one-to-one, it is a bijection. The proposition follows. □

Furthermore, it is not hard to see that X, Y disagree only on the colour assignments of the vertices in G_X. That is, the two variables give different colourings on the set Λ if and only if Λ and the vertex set of G_X intersect, that is,

$$\mathbb{P}[X(\Lambda) \neq Y(\Lambda)] = \mathbb{P}[\Lambda \cap G_X \neq \emptyset]. \tag{5.2}$$

Proof of Theorem 5.1. Let Ω_c and Ω_q be the set of k-colourings of G which assign vertex w colour c, q, respectively. Let X, Y be two random colourings such that $X(w) = c$ and $Y(w) = q$. Furthermore, assume that X, Y are coupled as described above. We have

$$\|\mu(\cdot \mid c) - \mu(\cdot \mid q)\|_\Lambda \leqslant \mathbb{P}[X(\Lambda) \neq Y(\Lambda)] = \mathbb{P}[\Lambda \cap G_X \neq \emptyset],$$

where the first inequality follows from the Coupling Lemma and the second equality is from (5.2).

The theorem follows by noting that we have $\Lambda \cap G_X \neq \emptyset$ only when there is at least one path from w to Λ such that X assigns to the vertices of this path colours c, q alternately. □

6. Application to $G(n, d/n)$

In this section we prove Theorem 2.1.

In the description of the counting schema in Section 4 we did not make any special assumptions as to how we construct $\mathcal{G}(G)$ on a given graph G. When we consider $G(n, d/n)$, we treat $\mathcal{G}(G(n, d/n))$ with a bit of care. We describe how we construct the sequence of subgraphs and we explain what motivates our choices.

Sequence of subgraphs $\mathcal{G}(G(n, d/n))$**.** Every cycle in $G(n, d/n)$ whose length is less than $\log n/(10 \log d)$ will be called *short*. Also, let $\mathcal{R}$ be the set that contains every edge in $G(n, d/n)$ which does not belong to a short cycle but one of its ends does.

In $\mathcal{G}(G(n, d/n))$, we choose G_0 to be empty. If i_0 is the number of edges in $G(n, d/n)$ that do not belong to $\mathcal{R}$, then for $i \leqslant i_0$, the edge e_{i-1}, in which G_i and G_{i-1} differ, is an arbitrary edge of $G(n, d/n)$ which is outside $\mathcal{R}$. On the other hand, for $i \geqslant i_0$ we have $e_i \in \mathcal{R}$.

https://doi.org/10.1017/S0963548320000255 Published online by Cambridge University Press

In $\mathcal{G}(G_i)$ the term $G_{i,0}$ is derived from G_{i,r_i} by deleting all the edges at distance $\log n/(20 \log d)$ from e_i, in arbitrary order.

In the following two paragraphs we explain what motivates the use of the above sequence of subgraphs. The discussion is a bit technical and the reader may skip it on a first reading of the paper.

We make use of the above sequence of subgraphs because at some point in the analysis we employ Lemma 4.3. Then, in order to meet the approximation guarantees, we have to ensure that the quantities $(\mathcal{C}_{i,j})^{-1}$ and $\mu_i(\mathcal{E}_i)$ are not too small. As it turns out, this could happen if we do not choose the sequence $\mathcal{G}(G(n, d/n))$ carefully enough. More specifically, we have to ensure that the two vertices in e_i or $e_{i,j}$ do not belong to the same short cycle which has large degree vertices, *i.e.* degree larger than k. The above construction of $\mathcal{G}(G(n, d/n))$ resolves exactly this problem.

In a typical instance of $G(n, d/n)$ the short cycles are far apart from each other. Removing the edges in $\mathcal{R}$ from the graph leaves the short cycles isolated from each other as well as isolated from the rest of the graph. This is the case for G_i, for all $i \leqslant i_0$.

As far as the terms G_i with $i > i_0$ are concerned, the estimation of $\mu_i(\mathcal{E}_i)$ is done by the trivial bound 1. Clearly this approximation is too rough. However, the error introduced into the estimation is very small due to the fact that the number of these terms, typically, is relatively small, *i.e.* $|\mathcal{R}| = O(n^{1/10} \log n)$.

Computing probabilities. The aim of this part is to argue that, for typical instances of $G(n, d/n)$, the computation of the quantities $\mu_{i,0}(\mathcal{E}_i)$ can be carried out in time which is polynomial in n. The main argument we use is that the component in $G_{i,0}$ which contains the vertices u_i, w_i is a very simple one: it is a tree with at most one extra edge. For such a simply structured component we employ dynamic programming (DP) to exactly compute $\mu_{i,0}(\mathcal{E}_i)$ in polynomial time. Recall that our objective is the exact computation of $\mu_{i,0}(\mathcal{E}_i)$.

First we argue that for every $e_i = \{u_i, w_i\}$ the graph that is induced by the vertices within distance $h = \log n/(20 \log d)$ from e_i is a tree with at most one extra edge.

Lemma 6.1. *Let $\mathcal{B}$ be the event that there is an edge e in $G(n, d/n)$ such that the graph that is induced by e and the vertices within distance $h = \log n/(20 \log d)$ from e is neither a tree nor a unicyclic graph. It holds that*

$$\mathbb{P}[\mathcal{B}] \leqslant n^{-3/4}.$$

Proof. We are going to show that, for any vertex e, the radius h neighbourhood cannot contain more than one cycle. If there is a neighbourhood with at least two 2 cycles, then it is not difficult to see that there is a set of $r \leqslant 4h$ vertices such that the number of edges exceeds the number of vertices by one. Then the lemma follows as a corollary from the following claim.

Claim 1. *With probability at least $1 - n^{-3/4}$, there is no subset of vertices of size at most $\log n/(5 \log d)$ in $G(n, d/n)$ which spans a number of edges that exceeds the number of vertices by at least one.*

Proof of Claim 1. Let $\ell = \log n/(5 \log d)$. Also, let D be the event that in $G(n, d/n)$ there exists a set of $r \leqslant \ell$ vertices which have at least $r + 1$ edges between them. Then we have

$$\begin{aligned}\mathbb{P}[D] &\leqslant \sum_{r=1}^{\ell} \binom{n}{r} \binom{\binom{r}{2}}{r+1} (d/n)^{r+1} \\ &\leqslant \sum_{r=1}^{\ell} \left(\frac{ne}{r}\right)^r \left(\frac{r^2 e}{2(r+1)}\right)^{r+1} (d/n)^{r+1}\end{aligned}$$

https://doi.org/10.1017/S0963548320000255 Published online by Cambridge University Press

$$\leqslant \frac{ed\ell}{2n}\sum_{r=1}^{\ell}\left(\frac{e^2d}{2}\right)^r$$

$$\leqslant \frac{ed\ell}{n}\left(\frac{e^2d}{2}\right)^{\ell}$$

$$\leqslant n^{-3/4}.$$

For the last inequality we recall that $\ell = \log n/(5\log d)$. The claim follows. □

The lemma follows. □

To see how can we compute the marginals, first consider a tree T and the list colourings of the tree. In the list colouring problem, every vertex u is associated with a $\texttt{List}(u)$ of colours. A list colouring of T is a proper colouring such that each u is assigned a colour from $\texttt{List}(u)$. It is clear that the list colouring is a generalization of the colouring. Using dynamic programming we can compute exactly the number of list colourings of a tree T. For a tree on N vertices, using dynamic programming we can compute exactly the number of list colourings in time Nk.

For $0 \leqslant i \leqslant i_0$, the connected component in $G_{i,0}$ that contains $\{w_i, u_i\}$ is a tree with at most one extra edge with probability that tends to 1 with n. If the component is a tree then we can count the number of colourings by working as indicated above. On the other hand, if the component is a tree with an extra edge, we can consider all the $k(k-1)$ colourings of the endpoints of the extra edge and for each of these colourings use dynamic programming to count the list colourings.

Using the above approach we can compute exactly the probability term $\mu_{i,0}(\mathcal{E}_i)$ by computing the number of relevant k-colourings of the connected component that $\{w_i, u_i\}$ belongs to. More specifically, we have to count the number of k-colourings which assign w_i, u_i different colours as well as the number of all the k-colourings of the component. Clearly $\mu_{i,0}(\mathcal{E}_i)$ follows by dividing these two quantities.

It is elementary to verify that we can compute *exactly* $\mu_{i,0}(\mathcal{E}_i)$ in time $O(k^4n)$, as the size of the component of each $\{w_i, u_i\}$ over which the computation takes place has size at most n.

The time required for the algorithm to compute all the probabilities $\mu_{i,0}(\mathcal{E}_i)$ is $O(k^4n)$ times r; recall that the number of terms the sequence $\mathcal{G}(G(n, d/n))$ is $r+1$. Since the number of terms is at most the number of edges in $G(n, d/n)$, a simple Chernoff bound implies that

$$\mathbb{P}[r \geqslant dn/2 + n^{2/3}] \leqslant \exp(-n^{1/5}).$$

From the above, we conclude the following result.

Corollary 6.2. *For any $d > 0$, with probability at least $1 - 2n^{-3/4}$ over the instances of $G(n, d/n)$, the counting schema requires $O(k^4n^2)$ steps to compute all the quantities $\mu_{i,0}(\mathcal{E}_i)$, for $i = 0, \ldots, r-1$.*

It could be that the aforementioned structural properties do not hold for the input graph, that is, the number of edges is more than $dn/2 + n^{2/3}$ or there are neighbourhoods which are not a tree with at most one extra edge. If this the case, the algorithm stops and returns an arbitrary number at the output. Furthermore, it is elementary to show that checking the aforementioned structural properties of the graph requires $O(n^2)$ time.

As far as the accuracy of the algorithm is concerned, we have the following result.

Proposition 6.3. *For $\varepsilon > 0$ there exists d_0 such that for $d > d_0$ the following is true. For integer $k \geqslant (1+\varepsilon)d$ and with probability at least $1 - 2n^{-\log(1+\varepsilon/4)/(52\log d)}$ over the instances $G(n, d/n)$, the counting schema computes an $(n^{-\log(1+\varepsilon/4)/(50\log d)})$-approximation of $\log Z(G(n, d/n), k)$.*

https://doi.org/10.1017/S0963548320000255 Published online by Cambridge University Press

The proof of Proposition 6.3 appears in Section 6.1 and uses Theorem 5.1.

A natural question at this point in the presentation is whether the algorithm can detect the cases of input graphs for which the error at the output is larger than that of the typical instances. The techniques we have introduced in Section 5 are only sufficient to *upper*-bound the errors we encounter in the analysis. In order to detect the untypical cases it seems that we need to have techniques that provide *lower* bounds for our errors. In that respect it does not seem to be a direct way for the algorithm to verify whether or not the output satisfies the approximation guarantees.

Theorem 2.1 follows as a corollary from the above results.

6.1 Proof of Proposition 6.3

Recall that $\mathcal{R}$ is the set of edges of $G(n, d/n)$ that do not belong to a cycle of length smaller than $\log n/(10 \log d)$, but they are incident to a vertex that belongs to such a cycle.

Lemma 6.4. *With probability at least* $1 - 3n^{-2/10}$ *over the instances of* $G(n, d/n)$*, the following events occur:*

A:= $|\mathcal{R}| < dn^{3/10}(\log n)$,

B:= *The number of cycles of length at most* $\log n/(10 \log d)$ *is less than* $2n^{3/10}$,

C:= *Removing* $\mathcal{R}$ *from* $G(n, d/n)$*, makes every cycle of length less than* $\log n/(10 \log d)$ *isolated.*

Proof. We start with **C**. The proof is very similar to that of Lemma 6.1.

Let $\delta = 1/(10 \log (d))$. Assume that after removing all the edges in $\mathcal{R}$ there are two cycles of length less than $\delta \log n$ which are connected, that is, these two cycles share vertices. Then there must exist a subgraph of $G(n, d/n)$ that contains at most $2\delta \log n$ vertices while the number of edges exceeds the number of vertices by one or more. From Claim 1 we have that the probability of having such a set in $G(n, d/n)$ is less than $n^{-3/4}$. This implies that the probability of having non-isolated short cycles in $G(n, d/n)$ once we remove the edges in $\mathcal{R}$ is less than $n^{-3/4}$, or

$$\mathbb{P}[\bar{\mathbf{C}}] \leqslant n^{-3/4}, \tag{6.1}$$

where $\bar{\mathbf{C}}$ is the complement of the event **C**. For some integer $\ell \geqslant 0$, let N_ℓ denote the set of cycles in $G(n, d/n)$ which are of length *at most* ℓ. It is elementary to show that $\mathbb{E}[|N_\ell|] \leqslant 2d^\ell$, which implies that $\mathbb{E}[|N_{\delta \log n}|] \leqslant 2n^{1/10}$. Also, we have $\mathbb{E}[|\mathcal{R}|] \leqslant dn^{1/10}(\log n)$. From Markov's inequality, we get the following two results:

$$\mathbb{P}[\bar{\mathbf{A}}] \leqslant \mathbb{P}\big[|\mathcal{R}| \geqslant n^{2/10}\mathbb{E}[|\mathcal{R}|]\big] \leqslant n^{-2/10}, \tag{6.2}$$

where $\bar{\mathbf{A}}$ is the complement of the event **A**. Also, we have

$$\mathbb{P}[\bar{\mathbf{B}}] \leqslant \mathbb{P}\big[|N_{\delta \log n}| \geqslant n^{2/10}\mathbb{E}[|N_{\delta \log n}|]\big] \leqslant n^{-2/10}. \tag{6.3}$$

The lemma follows from (6.1), (6.2), (6.3) and noting that

$$\mathbb{P}[\bar{\mathbf{A}} \cup \bar{\mathbf{B}} \cup \bar{\mathbf{C}}] \leqslant \mathbb{P}[\bar{\mathbf{A}}] + \mathbb{P}[\bar{\mathbf{B}}] + \mathbb{P}[\bar{\mathbf{C}}],$$

where the inequality follows from a simple union bound. □

We use the following theorem from [8].

Theorem 6.1 (Efthymiou [8])**.** *Let* $\varepsilon > 0$*, let* $d > 0$ *be sufficiently large and let fixed* $k \geqslant (1 + \varepsilon)d$*. Consider* $\boldsymbol{G} = G(n, d/n)$*. Let the graph* $\boldsymbol{H}$ *be such that* $V(\boldsymbol{H}) = V(\boldsymbol{G})$ *and* $E(\boldsymbol{H}) \subseteq E(\boldsymbol{G})$*. For any two* $c, q \in [k]$*, different from each other, any non-negative integer* $\ell \leqslant (\log n)^2$ *and a sequence* $P = (w_0, \ldots, w_\ell)$ *of vertices in* $V(\boldsymbol{H})$*, the following is true.*

https://doi.org/10.1017/S0963548320000255 Published online by Cambridge University Press

Let X be a random k-colouring of $\boldsymbol{H}$ conditional on $X(w_0) = c$. Let $\mathbf{I}_{\{P\}} = 1$ if P is a path in $\boldsymbol{H}$ and $X(w_i) \in \{c, q\}$, for every $j = 1, \dots, \ell$. Otherwise $\mathbf{I}_{\{P\}} = 0$. It holds that

$$\mathbb{P}[\mathbf{I}_{\{P\}} = 1] \leqslant 2[(1 + \varepsilon/4)n]^{-\ell}. \tag{6.4}$$

Remark 6.1. In (6.4) the probability term is with respect to both the randomness of the graph and the colouring X.

Furthermore, in [8] the following is noted. The above theorem implies that for $k \geqslant (1 + \varepsilon)d$, in a random k-colouring of $G(n, d/n)$, typically there are no long paths coloured with only two colours. This property is *monotone* in the graph structure. That is, it holds even if we remove an arbitrary number of edges from $G(n, d/n)$ (and get the subgraph $\boldsymbol{H}$).

For the graph $G = (V, E)$, let $w \in V$ and $\Lambda \subseteq V$ such that $w \notin \Lambda$. For X, a random k-colouring of G and for any integer $\ell \geqslant 0$ and $c, q \in [k]$, we define the random variable $\mathcal{D}_\ell(G, w, \Lambda, c, q)$ to be equal to the number of paths P from w to Λ which are of length at least ℓ, while for every $u \in P$ we have $X(u) \in \{c, q\}$. Also, we let

$$\mathcal{D}_\ell(G, w, \Lambda) = \max_{c,q \in [k]} \{\mathcal{D}_\ell(G, w, \Lambda, c, q)\}.$$

Using Theorem 6.1, we obtain the following result.

Proposition 6.5. *Let ε, d, k be as in Proposition 6.3. Consider $G(n, d/n)$ and let H, H' be arbitrary, spanning, subgraphs of $G(n, d/n)$ such that H is a subgraph of H'. For any vertex w, any subset of vertices Λ such that $w \notin \Lambda$, and for any positive integer $\ell \leqslant 100 \log n$, the following is true:*

$$\mathbb{P}[\mathcal{D}_\ell(H, w, \Lambda) > 0] \leqslant \mathbb{P}[\mathcal{D}_\ell(H', w, \Lambda) > 0] \leqslant 10k^2\varepsilon^{-1}n^{-1}\,|\Lambda|(1 + \varepsilon/4)^{-\ell}, \tag{6.5}$$

where $|\Lambda|$ is the cardinality of Λ.

For a proof of Proposition 6.5, see Section 6.2.

Lemma 6.6. *Let ε, d, k be as in Proposition 6.3. With probability at least $1 - n^{-\log(1+\varepsilon/4)/(22\log d)}$ over the instances of $G(n, d/n)$, the sequences $\mathcal{G}(G(n, d/n))$ and $\mathcal{G}(G_i)$, as defined in Section 6, have the following property. For any i such that e_i does not belong to a cycle of length less than $\log n/(10 \log d)$, we have*

$$\max_{\sigma,\eta \in [k]^{\{w_i\}}} \|\mu_i(\cdot \mid \sigma) - \mu_i(\cdot \mid \eta)\|_{u_i} \leqslant n^{-\log(1+\varepsilon/4)/(20\log d)},$$

and for any $i = 0, \dots, i_0$ and $j \in \{0, \dots, r_i + 1\}$, we have

$$\max_{\sigma,\eta \in [k]^{\{w_{i,j}\}}} \|\mu_{ij}(\cdot \mid \sigma) - \mu_{ij}(\cdot \mid \eta)\|_{u_{ij}} \leqslant n^{-\log(1+\varepsilon/4)/(20\log d)}.$$

Recall that the number of terms in $\mathcal{G}(G(n, d/n))$ is $r + 1$, while for every $i = 0, \dots, r$ the number of terms in $\mathcal{G}(G_i)$ is $r_i + 1$.

Proof. Let $\mathcal{L}$ be the set of edges that of $G(n, d/n)$ that do not belong to a cycle of length less than $\log n/(10 \log d)$. It is direct to check that the edges e_i and e_{ij} we consider here belong to $\mathcal{L}$.

For each edge $e \in \mathcal{L}$, we let $\mathcal{I}(e)$ be set of indices that appears in the algorithm, that is, either as an edge $e_{i,j}$ or as an edge e_i. Also, let $\mathcal{B}(e)$ be the event that either exists $(i, j) \in \mathcal{I}(e)$ such that

$$\max_{\sigma,\eta \in [k]^{\{w_{i,j}\}}} \|\mu_{ij}(\cdot \mid \sigma) - \mu_{ij}(\cdot \mid \eta)\|_{u_{ij}} > n^{-\log(1+\varepsilon/4)/(20\log d)},$$

https://doi.org/10.1017/S0963548320000255 Published online by Cambridge University Press

or $i \in \mathcal{I}(e)$ such that

$$\max_{\sigma,\eta \in [k]^{\{w_i\}}} \|\mu_i(\cdot \mid \sigma) - \mu_i(\cdot \mid \eta)\|_{u_i} > n^{-\log(1+\varepsilon/4)/(20 \log d)}.$$

For the sake of brevity, we refer to the elements of $\mathcal{I}(e)$ as singletons, that is, we use letters i or j rather than pairs of indices (i, j). To prove the lemma it suffices to show that

$$\mathbb{P}[\cup_{e \in \mathcal{L}} \mathcal{B}(e)] \leqslant n^{-\log(1+\varepsilon/4)/(22 \log d)}. \tag{6.6}$$

We prove (6.6) by showing that, for any edge e in $G(n, d/n)$,

$$\mathbb{P}[\mathcal{B}(e) \mid e \in \mathcal{L}] \leqslant (12k^2\varepsilon^{-1}) n^{-(1+\log(1+\varepsilon/4)/(20 \log d))}. \tag{6.7}$$

More specifically, we get (6.6) from (6.7) by working as follows:

$$\mathbb{P}[\cup_{e \in \mathcal{L}} \mathcal{B}(e)] \leqslant \sum_e \mathbb{P}[e \in \mathcal{L}] \mathbb{P}[\mathcal{B}(e) \mid e \in \mathcal{L}] \leqslant (12k^2\varepsilon^{-1}) n^{-(1+\log(1+\varepsilon/4)/(20 \log d))}\, \mathbb{E}[|\mathcal{L}|], \tag{6.8}$$

where the first summation runs over all possible pairs of vertices in $G(n, d/n)$. In the last inequality we use (6.7). Noting that $|\mathcal{L}|$ is at most the number of edges of $G(n, d/n)$, we have

$$\mathbb{E}[|\mathcal{L}|] \leqslant dn/2.$$

Plugging the above bound into (6.8) we get (6.6). It remains to show that (6.7) is indeed correct.

From Theorem 5.1 we have

$$\max_{\sigma,\eta \in [k]^{\{w_i\}}} \|\mu_i(\cdot \mid \sigma) - \mu_i(\cdot \mid \eta)\|_{u_i} \leqslant \mathbb{P}[\mathcal{D}_\ell(G_i, w_i, u_i) > 0],$$

where $\ell = \log n/(10 \log d)$. The probability term on the right-hand side is with respect to the Gibbs distribution on G_i. For every $e \in \mathcal{L}$, we let

$$\rho_e = \max_{i \in \mathcal{I}(e)} \{\mathbb{P}[\mathcal{D}_\ell(G_i, v_i, u_i) > 0]\}.$$

It is equivalent to the event $\mathcal{B}(e)$ to have

$$\rho_e \geqslant n^{-\log(1+\varepsilon/4)/(20 \log d)}.$$

From Proposition 6.5 we have $\rho_e \leqslant \mathbb{P}[\mathcal{D}_\ell(G_r \setminus \{e\}, v_i, u_i) > 0]$, where $G_r \setminus \{e\}$ is the initial graph minus the edge e.

For what follows, we consider a uniformly random edge $\mathbf{e} \in G(n, d/n)$. Let $\mathbf{x}, \mathbf{y}$ be the two ends of $\mathbf{e}$. We have that $\mathbf{x}, \mathbf{y}$ are two random vertices in $G(n, d/n)$. Let $\mathcal{A}$ be the event that $\mathbf{x}, \mathbf{y}$ are such that their distance is at least $\ell = \log n/(10 \log d)$, while none of them belongs to a short cycle. It is elementary to verify that $\mathbb{P}[\mathcal{A}] = 1 - o(1)$.

We consider the graph $\boldsymbol{G} = G(n, d/n) \setminus \{\mathbf{e}\}$, *i.e.* $G(n, d/n)$ without the random edge $\mathbf{e}$. From Proposition 6.5 we get the following. For $\hbar = n^{\log(1+\varepsilon/4)/(20 \log d)}$, we have

$$\begin{aligned}
\mathbb{P}[\mathcal{B}(e) \mid e \in \mathcal{L}] &\leqslant \mathbb{E}_{\boldsymbol{G}}[\mathbf{1}\{\mathbb{P}[\mathcal{D}_\ell(\mathbf{G}, \mathbf{x}, \mathbf{y}) > 0] \geqslant \hbar^{-1}\} \mid \mathcal{A}] && (6.9)\\
&\leqslant \mathbb{P}[\mathcal{A}]^{-1} \mathbb{E}_{\boldsymbol{G}}[\mathbf{1}\{\mathbb{P}[\mathcal{D}_\ell(\mathbf{G}, \mathbf{x}, \mathbf{y}) > 0] \geqslant \hbar^{-1}\}] \\
&\leqslant \mathbb{P}[\mathcal{A}]^{-1} \mathbb{E}_{\boldsymbol{G}}[\mathbf{1}\{\mathbb{E}_\mu[\mathcal{D}_\ell(\mathbf{G}, \mathbf{x}, \mathbf{y})] \geqslant \hbar^{-1}\}] && (6.10)\\
&\leqslant \hbar\, \mathbb{P}[\mathcal{A}]^{-1} \mathbb{E}_{\boldsymbol{G}}[\mathbb{E}_\mu[\mathcal{D}_\ell(\mathbf{G}, \mathbf{x}, \mathbf{y})]] && (6.11)\\
&\leqslant \hbar\, \mathbb{P}[\mathcal{A}]^{-1} \mathbb{E}[\mathcal{D}_\ell(\mathbf{G}, \mathbf{x}, \mathbf{y})]]. && (6.12)
\end{aligned}$$

https://doi.org/10.1017/S0963548320000255 Published online by Cambridge University Press

The expectation $\mathbb{E}_{\boldsymbol{G}}$ is with respect to the graph instances $\boldsymbol{G}$ only. In (6.9) the probability term inside the indicator is with respect to Gibbs distribution of $\boldsymbol{G}$. In (6.10) $\mathbb{E}_{\mu}[\,\cdot\,]$ is the expectation with respect to μ. The inequality in (6.11) follows from the observation that

$$\mathbf{1}\{\mathbb{E}_{\mu}[\mathcal{D}_{\ell}(\boldsymbol{G},\mathbf{x},\mathbf{y})] \geqslant \hbar^{-1}\} \leqslant \hbar\, \mathbb{E}_{\mu}[\mathcal{D}_{\ell}(\boldsymbol{G},\mathbf{x},\mathbf{y})]\}.$$

Using Proposition 6.5 for $\ell = \log n/(10\log d)$, we have

$$\mathbb{E}[\mathcal{D}_{\ell}(\boldsymbol{G},\mathbf{x},\mathbf{y})] \leqslant (10k^2\varepsilon^{-1})n^{-(1+\log(1+\varepsilon/4)/(10\log d))}. \tag{6.13}$$

Then (6.7) follows from (6.12) and (6.13) and the fact that $\mathbb{P}[\mathcal{A}] = 1 - o(1)$. The lemma follows. □

Using Lemma 6.6 we get two very useful corollaries.

Corollary 6.7. *Let ε, d, k be as in Proposition 6.3. For $\mathcal{G}(G(n,d/n))$, $\mathcal{G}(G_i)$ as defined in Section 6 and $\mathcal{C}_{i,j}$ as defined in Lemma 4.3, the following is true:*

$$\mathbb{P}[\mathcal{C}_{i,j} < 4k^2, \text{ for } 0 \leqslant i \leqslant i_0,\ 0 \leqslant j \leqslant r_i] \geqslant 1 - n^{-\log(1+\varepsilon/4)/(22\log d)}. \tag{6.14}$$

Proof. Let X be a random colouring of $G_{i,j} \in \mathcal{G}(G_i)$. Expressing $\mathcal{C}_{i,j}$ in terms of X, we have

$$\mathcal{C}_{ij} = 2 \max_{s,t\in[k]}\{(\mathbb{P}[X(u_{i,j}) = s \mid X(w_{i,j}) = t])^{-2}\}.$$

We deal with $\mathcal{C}_{i,j}$ by comparing $\mathbb{P}[X(u_{i,j}) = s \mid X(w_{i,j}) = t]$ with $\mathbb{P}[X(u_{i,j}) = s] = k^{-1}$. We show that these two probability terms are close to each other for any $s, t \in [k]$. In particular, we have

$$|\mathbb{P}[X(u_{i,j}) = s \mid X(w_{i,j}) = t] - \mathbb{P}[X(u_{i,j}) = s]| \leqslant \max_{\sigma,\eta\in[k]^{\{w_{i,j}\}}} \|\mu_{ij}(\cdot \mid \sigma) - \mu_{ij}(\cdot \mid \eta)\|_{u_{ij}}. \tag{6.15}$$

The result will follow by showing the following. With probability at least $1 - n^{-\log(1+\varepsilon/4)/(22\log d)}$ over the instances of $G(n,d/n)$, for all $0 \leqslant i \leqslant i_0$, $0 \leqslant j \leqslant r_i$ the left-hand side of (6.15) is smaller than k^{-2}. Clearly this is true due to Lemma 6.6. The corollary follows. □

Corollary 6.8. *Let ε, d, k be as in Proposition 6.3. For $\mathcal{G}(G(n,d/n))$ as in Section 6, the following holds.*

- *Let I contain every i such that e_i does not belong to a cycle of length less than $\log n/(10\log d)$. We have*

$$\mathbb{P}[\exists i \in I \text{ s.t. } |\mu_i(\mathcal{E}_i) - (1-k^{-1})| \geqslant n^{-\log(1+\varepsilon/4)/(20\log d)}] \leqslant n^{-\log(1+\varepsilon/4)/(22\log d)}. \tag{6.16}$$

- *Let I' contain every i such that e_i belongs to cycle of length less than $\log n/(10\log d)$. There is a constant $c > 0$, such that*

$$\mathbb{P}[\exists i \in I' \text{ s.t. } \mu_i(\mathcal{E}_i) < c] \leqslant 3n^{-2/10}. \tag{6.17}$$

Proof. We start with (6.16). Let X be a random k-colouring of G_i. We have

$$\begin{aligned}
|\mu_i(\mathcal{E}_i) - (1-k^{-1})| &= |\mathbb{P}[X(w_i) = X(u_i)] - k^{-1}| \\
&\leqslant \max_{\sigma\in[k]} |\mathbb{P}[X(u_i) = \sigma \mid X(w_i) = \sigma] - \mathbb{P}[X(u_i) = \sigma]| \\
&\leqslant \max_{\sigma,\eta\in[k]^{\{w_i\}}} \|\mu_i(\cdot \mid \sigma) - \mu_i(\cdot \mid \eta)\|_{u_i}.
\end{aligned} \tag{6.18}$$

https://doi.org/10.1017/S0963548320000255 Published online by Cambridge University Press

Equation (6.16) will follow once we show that with probability at least $1 - n^{-\log(1+\varepsilon/4)/(22\log d)}$ over the instances of $G(n, d/n)$, for all $i \in I$ the left-hand side of (6.18) is smaller than $n^{-\log(1+\varepsilon/4)/(20\log d)}$. Clearly this is true due to Lemma 6.6.

In the latter case, *i.e.* (6.17), we consider w_i and u_i which belong to a short cycle. Such a pair of vertices appears in the schema only when we have removed from $G(n, d/n)$ all the edges in $\mathcal{R}$. By Lemma 6.4 we have that with probability at least $1 - 3n^{-2/10}$ the removal of the edges in $\mathcal{R}$ disconnects every small cycle from the rest of $G(n, d/n)$. Thus, for the second case, where v_i, u_i belong to a small isolated cycle, $\mathbb{P}[X(u_i) \neq X(w_i)]$ is trivially lower-bounded by some constant $c > 0$, since $k > 2$.

The corollary follows. □

We have all the results we need to prove Proposition 6.3.

Proof of Proposition 6.3. Consider $\mathcal{G}(G(n, d/n))$, and for each $0 \leqslant i \leqslant r$ consider $\mathcal{G}(G_i)$. Recall that $r + 1$ is the number of terms in $\mathcal{G}(G(n, d/n))$, while $r_i + 1$ is the number of terms in $\mathcal{G}(G_i)$. Also, recall that μ_i is the Gibbs distribution of G_i, while $\mathcal{R}$ contains every edge in $G(n, d/n)$ which does not belong to a short cycle but one of its ends does.

Let $\mathcal{W}$ be the event that the following sub-events occur:

$\mathcal{W}_1$: $|\mathcal{R}| \leqslant dn^{3/10} \log n$,

$\mathcal{W}_2$: $\min_{0\leqslant i\leqslant r}\{\mu_i(\mathcal{E}_i)\} = \Theta(1)$ and for any $i \leqslant i_0$ we have $\max_j (\mathcal{C}_{i,j}) \leqslant 4k^2$,

$\mathcal{W}_3$: $r \leqslant dn/2 + n^{2/3}$ and $\max_{i\leqslant i_0}\{r_i\} \leqslant n^{1/20}(\log n)^3$.

Claim 2. *It holds that* $\mathbb{P}[\mathcal{W}] \geqslant 1 - n^{-\log(1+\varepsilon/4)/(25\log d)}$.

The proof of Claim 2 appears at the end of this proof. We also need the following claim.

Claim 3. *For* $i \leqslant i_0$, *we have*

$$\mathbb{E}[|\mu_i(\mathcal{E}_i) - \mu_{i,0}(\mathcal{E}_i)| \mid \mathcal{W}] \leqslant n^{-\log(1+\varepsilon/4)/(25\log d)}.$$

The proof of Claim 3 appears after this proof.

Let $\mathcal{Z}$ be the estimate of the algorithm for $Z(G(n, d/n), k)$. If the event $\mathcal{W}$ holds, then the following hold as well. For any $0 \leqslant i \leqslant r$, we have

$$\frac{|\mu_i(\mathcal{E}_i) - \mu_{i,0}(\mathcal{E}_i)|}{\mu_i(\mathcal{E}_i)} = O(1) \quad \text{and} \quad \frac{|\mu_i(\mathcal{E}_i) - \mu_{i,0}(\mathcal{E}_i)|}{\mu_{i,0}(\mathcal{E}_i)} = O(1). \tag{6.19}$$

Recall that for every $i > i_0$ the schema estimates $\mu_i(\mathcal{E}_i)$ to be equal to 1.

Combining (6.19), Claim 3 and Proposition 4.2, we get

$$\mathbb{E}[n^{-1}|\log \mathcal{Z} - \log Z(G(n, d/n), k)| \mid \mathcal{W}] \leqslant 2dn^{-\log(1+\varepsilon/4)/(25\log d)}.$$

Using Markov's inequality, from the above we get

$$\mathbb{P}[n^{-1}|\log \mathcal{Z} - \log Z(G(n, d/n), k)| \geqslant n^{-\log(1+\varepsilon/4)/(50\log d)} \mid \mathcal{W}] \leqslant n^{-\log(1+\varepsilon/4)/(52\log d)}.$$

The proposition follows from the above inequality and the fact that

$$\mathbb{P}[\mathcal{W}] \geqslant 1 - n^{-\log(1+\varepsilon/4)/(25\log d)}.$$ □

Proof of Claim 2. In light of Lemma 6.4, we immediately have $\mathbb{P}[\bar{\mathcal{W}}_1] \leqslant 3n^{-2/10}$. Furthermore, from Corollaries 6.8 and 6.7 we have

$$\mathbb{P}[\bar{\mathcal{W}}_2] \leqslant 2n^{-\log(1+\varepsilon/4)/(22\log d)}.$$

https://doi.org/10.1017/S0963548320000255 Published online by Cambridge University Press

We are going to show that

$$\mathbb{P}[\bar{\mathcal{W}}_3] \leqslant n^{-1}. \tag{6.20}$$

Then the claim follows by combining the above bounds with

$$\mathbb{P}[\bar{\mathcal{W}}] = \mathbb{P}[\bar{\mathcal{W}}_1 \cup \bar{\mathcal{W}}_2 \cup \bar{\mathcal{W}}_3] \leqslant \mathbb{P}[\bar{\mathcal{W}}_1] + \mathbb{P}[\bar{\mathcal{W}}_2] + \mathbb{P}[\bar{\mathcal{W}}_3],$$

where the last inequality is the union bound.

Let $\mathcal{J}_1$ be the event than $r \geqslant dn/2 + n^{2/3}$. The number of terms in $\mathcal{G}(n, d/n)$ is upper-bounded by the number of edges in the graph. Applying Chernoff's bound for the number of edges of $G(n, d/n)$, we obtain

$$\mathbb{P}[\mathcal{J}_1] \leqslant \exp\big(-n^{1/5}\big). \tag{6.21}$$

Let $\mathcal{H}_i$ be the event that $r_i > n^{1/20}(\log n)^3$, and let $\mathcal{J}_2 = \cup_i \mathcal{H}_i$. In light of (6.21), it suffices to show that

$$\mathbb{P}[\mathcal{J}_2] \leqslant n^{-3/2}, \tag{6.22}$$

since a simple union bound implies that $\mathbb{P}[\bar{\mathcal{W}}_3] \leqslant \mathbb{P}[\mathcal{J}_1] + \mathbb{P}[\mathcal{J}_2]$.

Let $\hbar = \log n/(20 \log d)$. We note that r_i is equal to L, the set of vertices at distance $\hbar + 2$ from e_i. The size of L is dominated by the number of vertices at level $\hbar + 1$ of a Galton–Watson tree, with the number of offspring per individual being $\mathcal{B}(n, d/n)$ and initial population 2. With standard arguments (see *e.g.* Theorem 6 in [18]), it holds that

$$\mathbb{P}\left[|L| > n^{1/20}(\log n)^3\right] \leqslant n^{-3}. \tag{6.23}$$

Then we have

$$\mathbb{P}[\mathcal{H}_i] \leqslant \mathbb{P}\left[|L| > n^{1/20}(\log n)^3\right] \leqslant n^{-3}. \tag{6.24}$$

Also, we have

$$\begin{aligned}
\mathbb{P}[\mathcal{J}_2] &= \mathbb{P}[\cup_i \mathcal{H}_i] \\
&\leqslant \mathbb{P}[\cup_i \mathcal{H}_i \mid \bar{\mathcal{J}}_1] + \mathbb{P}[\mathcal{J}_1] \\
&\leqslant \mathbb{P}[\cup_i \mathcal{H}_i \mid \bar{\mathcal{J}}_1] + \exp\left(-n^{1/5}\right) && \text{(from (6.21))} \\
&\leqslant \left(dn/2 + n^{2/3}\right) \mathbb{P}[\mathcal{H}_i \mid \bar{\mathcal{J}}_1] + \exp\left(-n^{15}\right) && \text{(union bound)} \\
&\leqslant \left(dn/2 + n^{2/3}\right) \frac{\mathbb{P}[\mathcal{H}_i]}{\mathbb{P}[\bar{\mathcal{J}}_1]} + \exp\left(-n^{1/5}\right) \\
&\leqslant n^{-3/2} && \text{(from (6.24) and (6.21)).}
\end{aligned}$$

The above proves (6.22). The claim follows. □

Proof of Claim 3. Letting

$$Q_{i,j} = \max_{\sigma,\tau \in [k]^{\{v_{ij}\}}} \{\|\mu_{i,j}(\cdot\,|\sigma) - \mu_{ij}(\cdot\,|\tau)\|_{e_i \cup \{u_{i,j}\}} + \|\mu_{i,j}(\cdot\,|\sigma) - \mu_{ij}(\cdot\,|\tau)\|_{u_{ij}}\},$$

Lemma 4.3 implies that for every i we have

$$\mathbb{E}[|\mu_{i,0}(\mathcal{E}_i) - \mu_i(\mathcal{E}_i)| \mid \mathcal{W}] \leqslant \mathbb{E}\left[\sum_{j=0}^{r_i-1} \mathcal{C}_{i,j}\, Q_{ij} \mid \mathcal{W}\right] \leqslant 4k^2\, \mathbb{E}\left[\sum_{j=0}^{r_i-1} Q_{i,j} \mid \mathcal{W}\right], \tag{6.25}$$

where the last derivation holds since, conditioning on $\mathcal{W}$, we have $\mathcal{C}_{ij} \leqslant 4k^2$. First we note that

$$\mathbb{E}[Q_{ij} \mid \mathcal{W}] \leqslant \frac{\mathbb{E}[Q_{ij}]}{\mathbb{P}[\mathcal{W}]} \leqslant \frac{4}{3}\mathbb{E}[Q_{ij}] \quad (\text{as } \mathbb{P}[\mathcal{W}] > 3/4). \tag{6.26}$$

https://doi.org/10.1017/S0963548320000255 Published online by Cambridge University Press

For $i \leqslant i_0$ and any $j = 0, \ldots, r_i$, we have

$$\mathbb{E}[Q_{ij}] \leqslant n^{-1/20 - \log(1+\varepsilon/4)/(23 \log d)}. \tag{6.27}$$

We get the above bound by working in a very similar way to Lemma 6.6. For this reason we omit the derivations. Plugging (6.26) into (6.25) and noting that, conditional on $\mathcal{W}$, we have $r_i \leqslant n^{1/20}(\log n)^3$, for any $i \leqslant i_0$ we obtain

$$\mathbb{E}[|\mu_{i,0}(\mathcal{E}_i) - \mu_i(\mathcal{E}_i)| \mid \mathcal{W}] \leqslant n^{-\log(1+\varepsilon/4)/(25 \log(d))}.$$

The claim follows. □

6.2 Proof of Proposition 6.5

Proof of Proposition 6.5. Consider an enumeration of all the sequences of t vertices in H with first the vertex w and last some vertex in Λ. Let $\pi_0(t), \pi_1(t), \ldots$ be the sequences in the order they appear in the enumeration. Also consider an enumeration of all sequences of exactly

$$\ell_0 = \frac{(\log n)^2}{\log(1+\varepsilon/4)}$$

vertices such that the first vertex is w while the last vertex can be *any* vertex of the graph, *i.e.* not necessarily in Λ. Let $\delta_0, \delta_1, \ldots$ be the sequences in the order they appear in the enumeration.

Consider X, a random k-colouring of H, that is, X is distributed as in $\mu = \mu_{H,k}$. Fix $c, q \in [k]$ such that $c \neq q$. For the sequence $\pi = x_0, \ldots, x_\ell$, we let the variable $\mathcal{Z}(\pi) = 1$ if the path $x_0, x_1, \ldots, x_\ell$ exists in H and $X(x_i) \in \{c, q\}$, for every $i \in \{0, 1, \ldots, s\}$. Otherwise $\mathcal{Z}(\pi) = 0$.

Let

$$\mathcal{Z}(t) = \sum_{i \geqslant 0} \mathcal{Z}(\pi_i(t)) \quad \text{and} \quad \mathcal{H} = \sum_{i \geqslant 0} \mathcal{Z}(\delta_i).$$

Also, let $\mathcal{E}$ be the event that 'in X there is a path P that starts from w and has length at least ℓ_0 such that for all $u \in P$ we have $X(u) \in \{c, q\}$'. We have

$$\begin{aligned}
\mathbb{P}\Big[\sum_{t \geqslant \ell} \mathcal{Z}(t) > 0\Big] &\leqslant \mathbb{P}\Big[\sum_{t \geqslant \ell} \mathcal{Z}(t) > 0, \bar{\mathcal{E}}\Big] + \mathbb{P}[\mathcal{E}] \\
&\leqslant \mathbb{P}\Big[\sum_{\ell \leqslant t < \ell_0} \mathcal{Z}(t) > 0\Big] + \mathbb{P}[\mathcal{E}] \\
&= \mathbb{P}\Big[\sum_{\ell \leqslant t < \ell_0} \mathcal{Z}(t) > 0\Big] + \mathbb{P}[\mathcal{H} > 0].
\end{aligned} \tag{6.28}$$

For the last inequality we use the observation that the event $\mathcal{E}$ occurs if and only if $\mathcal{H} > 0$.

Recall now that H is a subgraph of $G(n, d/n)$. Averaging over the base graph $G(n, d/n)$, (6.28) yields

$$\mathbb{P}[\mathcal{D}_\ell(H, w, \Lambda, c, q) > 0] \leqslant \mathbb{P}\Big[\sum_{\ell \leqslant t < \ell_0} \mathcal{Z}(t) > 0\Big] + \mathbb{P}[\mathcal{H} > 0]. \tag{6.29}$$

Note that in (6.28) the probability terms are with respect to the measure $\mu_{H,k}$, while in (6.29) the probabilities are with respect to both μ and the measure over the graph instances induced by

https://doi.org/10.1017/S0963548320000255 Published online by Cambridge University Press

$G(n, d/n)$. Markov's inequality implies that

$$\mathbb{P}\Big[\sum_{\ell\leqslant t<\ell_0}\mathcal{Z}(t)>0\Big]\leqslant\mathbb{E}\Big[\sum_{\ell\leqslant t<\ell_0}\mathcal{Z}(t)\Big]\leqslant\sum_{\ell\leqslant t<\ell_0}\mathbb{E}[\mathcal{Z}(t)]\leqslant 2|\Lambda|\,n^{-1}\sum_{\ell\leqslant t<\ell_0}((1+\varepsilon/4))^{-t},\tag{6.30}$$

where in the second inequality we use the linearity of expectation, while in the last inequality we use Theorem 6.1 and the fact that between x_1 and Λ there are at most $|\Lambda|\cdot n^{t-1}$ paths of length exactly t. With elementary calculations, the above yields

$$\mathbb{P}\Big[\sum_{\ell\leqslant t<\ell_0}\mathcal{Z}(t)>0\Big]\leqslant(8/\varepsilon)\,|\Lambda|\,n^{-1}(1+\varepsilon/4)^{-\ell}.\tag{6.31}$$

Also, Markov's inequality implies that

$$\begin{aligned}\mathbb{P}[\mathcal{H}>0]\leqslant\mathbb{E}[\mathcal{H}]&\leqslant 2n^{\ell_0}((1+\varepsilon/4)n)^{-\ell_0}\quad&&\text{(from Theorem 6.1)}\\&\leqslant 2n^{-(\log n)}&&\text{(since }\ell_0=\Theta((\log n)^2)).\end{aligned}\tag{6.32}$$

Since we assumed that $\ell=\Theta(\log n)$, we obtain

$$\mathbb{P}\Big[\sum_{\ell\leqslant t<\ell_0}\mathcal{Z}(t)>0\Big]\gg\mathbb{P}[\mathcal{H}>0].$$

Then, plugging (6.31) and (6.32) into (6.29), we get

$$\mathbb{P}[\mathcal{D}_\ell(H,w,\Lambda,c,q)>0]\leqslant n^{-1}(10/\varepsilon)\,|\Lambda|(1+\varepsilon/4)^{-\ell}.$$

The bound for $\mathbb{P}[\mathcal{D}_\ell(H,w,\Lambda)>0]$ follows by taking a union bound over the $k(k-1)$ pairs of of colours $c,q\in[k]$.

Using exactly the same line of arguments we get the same bound for $\mathbb{P}[\mathcal{D}_\ell(H',w,\Lambda)>0]$. The first inequality in the statement follows from the monotonicity of the event $\mathcal{D}_\ell(H,w,\Lambda)$.

The proposition follows. □

7. Proof of Theorem 2.2

Let $w_1,\dots,w_\ell$ be the vertices in Λ, where $\ell=|\Lambda|$. Recall that we have assumed that $\ell=o(\log n)$.

Without loss of generality assume that the indices of the vertices have the following property. The vertices with smaller indices are closer to w_1, that is, there are no two indices s,t such that $s>t$ and $\mathtt{dist}(w_1,w_s)<\mathtt{dist}(w_1,w_t)$.

Let μ be the Gibbs distribution over the k-colourings of $G(n,d/n)$. For each $j=1,\dots,\ell$, let $\Lambda_j=\{w_1,\dots,w_j\}$, while $\Lambda_0=\emptyset$. For any $\sigma\in[k]^\Lambda$, we have

$$\mu(\sigma)=\prod_{j=1}^{\ell}\mu(\sigma(w_j)\,|\,\sigma(\Lambda_{j-1})).\tag{7.1}$$

The algorithm we introduce aims to compute each term $\mu(\sigma(w_j)\,|\,\sigma(\Lambda_{j-1}))$ above. However, instead of computing these probability terms exactly, the algorithm approximates them. The challenge is to compute each of these probability terms sufficiently accurately and efficiently.

The approach is not too different from the one we use for the counting algorithm in Theorem 2.1. Instead of computing the aforementioned probability terms with respect to the graph $G(n,d/n)$, they are computed with respect to another, simpler, graph. Roughly speaking, the simpler graph is derived from $G(n,d/n)$ by deleting all the edges at some distance h from Λ, where $h=\log n/(7\log d)$.

https://doi.org/10.1017/S0963548320000255 Published online by Cambridge University Press

Let us be more specific. Let R_1 be the set of edges at distance $h = \log n/(7 \log d)$ from the set Λ. Let $\mathcal{C}$ be the set of cycles C of length at most $\ell_0 = \log n/(10(\log d)^2)$ and have edges which belongs to R_1. Let R_2 be the set of edges which have exactly one end in a cycle $C \in \mathcal{C}$.

Consider the sequence $\mathcal{G}(G(n, d/n))$ of $r+1$ terms, such that G_r is identical to $G(n, d/n)$ and G_0 is derived from $G(n, d/n)$ by deleting all the edges in $R_1 \cup R_2$. As we consider the sequence $G_0, \ldots, G_r$, first we insert the edges in $R_1 \setminus R_2$ and then the edges in R_2. That is, for $i \leqslant |R_1 \setminus R_2|$ we have that G_i and G_{i-1} differ only on the edge $e_{i-1} \in R_1$. For $i > |R_1 \setminus R_2|$, G_i and G_{i-1} differ only on the edge in $e_{i-1} \in R_2$.

The choice of the above sequence is made to avoid a situation where the edge e_i belongs to a cycle of length at most ℓ_0 and the vertices in this cycle are of high degree, *i.e.* larger than k. Recall that we have a similar situation for the counting algorithm. For typical instances of $G(n, d/n)$, removing the edges in R_2 isolates the cycles of length $\leqslant \ell_0$ from the rest of the graph and from each other. Our choice of $\mathcal{G}(G(n, d/n))$ guarantees that even though for some $i \leqslant |R_1 \setminus R_2|$ the edge e_i belongs to a cycle in $\mathcal{C}$, this cycle is isolated in the corresponding graph.

Let μ_i be the Gibbs distribution over the k-colourings of $G_i \in \mathcal{G}(G(n, d/n))$. For the sake of brevity we let ν denote the Gibbs distribution over the k-colourings of G_0. The algorithm uses $\nu(\sigma(w_j) \mid \sigma(\Lambda_{j-1}))$ to approximates $\mu(\sigma(w_j) \mid \sigma(\Lambda_{j-1}))$, for $j = 1, \ldots, \ell$.

For the typical instances of $G(n, d/n)$, the connected component in G_0 that contains w_j is a tree with at most one extra edge, this follows by working as in Lemma 6.1. This also implies that the computation of $\nu(\sigma(w_j) \mid \sigma(\Lambda_{j-1}))$ can be carried out by using dynamic programming, as we described in Section 6. Recall that the time required for the dynamic program to compute $\nu(\sigma(w_j) \mid \sigma(\Lambda_{j-1}))$ is $O(k^3 n)$. Then we conclude the following corollary.

Corollary 7.1. *With probability $1 - n^{-3/4}$ over the instances of $G(n, d/n)$, the estimation of marginals requires $O(|\Lambda| k^3 n)$ steps to compute all the quantities $\nu(\sigma(w_i) \mid \sigma(\Lambda_{i-1}))$, for $i = 1, \ldots, |\Lambda|$.*

For estimating the marginals we require the input graph to be such that the neighbourhood of the vertices $w_1, w_2, \ldots, w_\ell$ is a tree with at most one extra edge. If this not the case, the algorithm stops and returns an arbitrary number at the output. Also, it is elementary to show that checking these structural properties of the graph requires $O(n)$ time.

Theorem 2.2 follows by showing that $\nu(\sigma(w_j) \mid \sigma(\Lambda_{j-1}))$ and $\mu(\sigma(w_j) \mid \sigma(\Lambda_{j-1}))$ are not too far from each other. In particular, we have the following result.

Proposition 7.2. *For $\varepsilon > 0$, there exists $d_0 = d_0(\varepsilon)$ such that for $d > d_0$, for integer $k \geqslant (1+\varepsilon)d$ and with probability at least $1 - 2n^{-\log(1+\varepsilon/4)/(19 \log d)}$ over the instances of $G(n, d/n)$, the following holds. For all $j = 1, \ldots, |\Lambda|$ and for any $\sigma \in [k]^\Lambda$, we have*

$$|\nu(\sigma(w_j) \mid \sigma(\Lambda_{j-1})) - \mu(\sigma(w_j) \mid \sigma(\Lambda_{j-1}))| \leqslant n^{-\log(1+\varepsilon/4)/(19 \log d)}.$$

The proof of Proposition 7.2 appears in Section 7.1.

Let $\mathcal{A}_i$ be the event that

$$|\nu(\sigma(w_i) \mid \sigma(\Lambda_{i-1})) - \mu(\sigma(w_i) \mid \sigma(\Lambda_{i-1}))| \leqslant n^{-\log(1+\varepsilon/4)/(19 \log d)}.$$

It is elementary to verify that if $\cap_{i=1}^{\ell} \mathcal{A}_i$ holds, then for any $\sigma \in [k]^\Lambda$ we have

$$|\mu(\sigma) - \nu(\sigma)| = \left| \mu(\sigma) - \prod_{i=1}^{\ell} \nu(\sigma(w_i) \mid \sigma(\Lambda_{i-1})) \right| \leqslant (\log n)^2 n^{-\log(1+\varepsilon/4)/(19 \log d)}.$$

Combining the above with the definition of total variation distance, we obtain

$$\|\mu_\Lambda - \nu\| \leqslant k^{|\Lambda|} (\log n)^2 n^{-\log(1+\varepsilon/4)/(19 \log d)} \leqslant n^{-\log(1+\varepsilon/4)/(20 \log d)},$$

https://doi.org/10.1017/S0963548320000255 Published online by Cambridge University Press

where in the second inequality we use that $|\Lambda| = o(\log n)$.

The theorem follows by computing an appropriate bound of the probability of the event $\cap_i \mathcal{A}_i$. Note that Proposition 7.2 implies that

$$\mathbb{P}[\cup_i \bar{\mathcal{A}}_i] \leqslant 2n^{-\log(1+\varepsilon/2)/(19\log d)}.$$

This concludes the proof of Theorem 2.2. □

A natural question at this point is whether the algorithm can detect the cases of input graphs for which the error at the output is larger than the typical one. As in the case of Theorem 2.1, there does not seem to be a direct way for the algorithm to verify whether or not the output satisfies the approximation guarantees.

7.1 Proof of Proposition 7.2

The proof of Proposition 7.2 uses derivations which are very similar to those that appear for the proof of Theorem 2.1. To prove the proposition we need to use the following two results.

Lemma 7.3. *In the context of Proposition 7.2, with probability at least* $1 - n^{-\log(1+\varepsilon/4)/(19\log d)}$ *over the graph instances* $G(n, d/n)$*, the following is true. For all* $j = 1, \ldots, |\Lambda|$ *we have*

$$\|\mu - \nu\|_{\Lambda_j} \leqslant n^{-\log(1+\varepsilon/4)/(18\log d)}. \tag{7.2}$$

The proof of Lemma 7.3 appears in Section 7.2.

Furthermore, we are going to show the following.

Claim 4. *In the context of Proposition 7.2, with probability* $1 - n^{-3/4}$ *over the graph instances* $G(n, d/n)$*, the following is true. There exists a positive fixed number* $C < 1$ *such that, for any* $j = 1, \ldots, |\Lambda|$*, either* $\nu(\sigma(\Lambda_j)) = 0$*, or*

$$\nu(\sigma(\Lambda_j)) \geqslant C^j \geqslant C^{|\Lambda|} = \Omega(n^{-1/(\log d)^{20}}). \tag{7.3}$$

Proof. We assume that in G_0 the component that contains Λ_i is a tree with at most one edge. Working as in Claim 1, we get that this holds with probability at least $1 - n^{-3/4}$.

Then we use the assumption that for all w_t, with $t < j$ we have $\texttt{dist}(w_1, w_t) \leqslant \texttt{dist}(w_0, w_j)$. Since the component that contains Λ_j is a tree with at most one vertex, the following holds. Conditioning on every $w_1, \ldots w_{j-1}$, we cannot restrict the colouring of more than two neighbours of w_j. Specifically, if the component is tree-like, then $w_1, \ldots w_{j-1}$ cannot restrict the colouring of more than one neighbour of w_j. On the other hand, it is easy to check that for the unicyclic case the number of neighbours can be at most two.

The above implies that whatever the colourings of $w_1, \ldots w_{j-1}$, the probability of w_j getting colour $\sigma(w_j)$ in a random k-colouring of G_0 is either zero or at least C, for a fixed $C \in (0, 1)$. The first inequality in (7.3) follows by simple induction on j. The second inequality in (7.3) follows since $j \leqslant |\Lambda|$ and $C \in (0, 1)$. The last one follows from the fact that $|\Lambda| = o(\log n)$. □

Proof of Proposition 7.2. Assume that $G(n, d/n)$ satisfies (7.2) and (7.3). For all $j = 1, \ldots, |\Lambda|$ we have

$$|\nu(\sigma(\Lambda_j)) - \mu(\sigma(\Lambda_j))| \leqslant \|\mu - \nu\|_{\Lambda_j}. \tag{7.4}$$

Furthermore, since

$$\nu(\sigma(w_j) \mid \sigma(\Lambda_{j-1})) = \frac{\nu(\sigma(\Lambda_j))}{\nu(\sigma(\Lambda_{j-1}))} \quad \text{and} \quad \mu(\sigma(w_j) \mid \sigma(\Lambda_{j-1})) = \frac{\mu(\sigma(\Lambda_j))}{\mu(\sigma(\Lambda_{j-1}))},$$

https://doi.org/10.1017/S0963548320000255 Published online by Cambridge University Press

we have

$$\begin{aligned}
&|\nu(\sigma(w_j) \mid \sigma(\Lambda_{j-1})) - \mu(\sigma(w_j) \mid \sigma(\Lambda_{j-1}))| \\
&\leqslant \left| \frac{\nu(\sigma(\Lambda_j)}{\nu(\sigma(\Lambda_{j-1})} - \frac{\mu(\sigma(\Lambda_j)}{\mu(\sigma(\Lambda_{j-1}))} \right| \\
&\leqslant (\nu(\sigma(\Lambda_{j-1})))^{-1} \left| \nu(\sigma(\Lambda_j)) - \mu(\sigma(\Lambda_j)) \frac{\nu(\sigma(\Lambda_{j-1}))}{\mu(\sigma(\Lambda_{j-1}))} \right| \\
&\leqslant (\nu(\sigma(\Lambda_{j-1})))^{-1} \left| \nu(\sigma(\Lambda_j) - \mu(\sigma(\Lambda_j)) \left(1 + \frac{\nu(\sigma(\Lambda_{j-1})) - \mu(\sigma(\Lambda_{j-1}))}{\mu(\sigma(\Lambda_{j-1}))} \right) \right| \\
&\leqslant \frac{|\nu(\sigma(\Lambda_{j-1})) - \mu(\sigma(\Lambda_{j-1}))|}{\nu(\sigma(\Lambda_{j-1}))} + \frac{\mu(\sigma(\Lambda_j))}{\mu(\sigma(\Lambda_{j-1}))} \frac{|\nu(\sigma(\Lambda_j)) - \mu(\sigma(\Lambda_j))|}{\nu(\sigma(\Lambda_{j-1}))} \\
&\leqslant \frac{|\nu(\sigma(\Lambda_{j-1})) - \mu(\sigma(\Lambda_{j-1}))|}{\nu(\sigma(\Lambda_{j-1}))} + \frac{|\nu(\sigma(\Lambda_j)) - \mu(\sigma(\Lambda_j))|}{\nu(\sigma(\Lambda_{j-1}))},
\end{aligned}$$

where the last inequality follows from the observation that $\mu(\sigma(\Lambda_j))/\mu(\sigma(\Lambda_{j-1})) \leqslant 1$. Using (7.4), we obtain

$$\begin{aligned}
|\nu(\sigma(w_j) \mid \sigma(\Lambda_{j-1})) - \mu(\sigma(w_j) \mid \sigma(\Lambda_{j-1}))| &\leqslant \frac{\|\mu - \nu\|_{\Lambda_{j-1}} + \|\mu - \nu\|_{\Lambda_j}}{\nu(\sigma(\Lambda_{j-1}))} \\
&\leqslant 2 \frac{n^{-\log(1+\varepsilon/4)/(18 \log d)}}{\nu(\sigma(\Lambda_{j-1}))} \qquad \text{(from Lemma 7.3)} \\
&\leqslant n^{-\log(1+\varepsilon/4)/(19 \log d)} \qquad \text{(from Claim 4).}
\end{aligned}$$

Recall that the above holds if $G(n, d/n)$ satisfies (7.2) and (7.3). The proposition follows by noting that both (7.2) and (7.3) hold with probability at least $1 - 2n^{-\log(1+\varepsilon/4)/(19 \log d)}$. □

7.2 Proof of Lemma 7.3

Let $\hbar$ be the maximum distance between w_1 and any vertex in Λ. Recall that the assumption is that $\hbar = o(\log n)$. For the sake of definiteness let $\hbar = (\log n)/\omega(1)$, where $\omega(1)$ is an arbitrarily growing function of n. Set $\mathcal{S} = \mathtt{Sphere}_{\hbar}(w_1)$. Note that $\Lambda \subseteq \mathcal{S}$.

Recall that $r + 1$ is the number of terms in $\mathcal{G}(G(n, d/n))$, while for any G_i, G_{i+1} we have that the second graph is obtained from the first one by adding the edge $e_i = \{v_i, u_i\}$.

Let $\mathcal{D}_1$ be the event that $r \geqslant N = n^{1/7}(\log n)^4$. Also, let $\mathcal{D}_2$ be the event that $|\mathcal{S}| \geqslant n^{1/\omega(1)}$, while there are two cycles in $G(n, d/n)$ which are of length at most $\ell_0 = \log n/(10(\log d)^2)$ and share vertices. Finally, let $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$.

Working as in the proof of Proposition 6.3 (*e.g.* Lemma 6.4 and Claim 2), we obtain

$$\mathbb{P}[\mathcal{D}] \leqslant n^{-1/2}. \tag{7.5}$$

Furthermore, for a given j, we have

$$\begin{aligned}
\mathbb{E}[\|\mu - \nu\|_{\Lambda_j}] &\leqslant \mathbb{E}[\|\mu - \nu\|_{\mathcal{S}}] \\
&\leqslant \mathbb{E}[\|\mu - \nu\|_{\mathcal{S}} \mid \bar{\mathcal{D}}] + \mathbb{P}[\mathcal{D}] \qquad \text{(since } \|\mu - \nu\|_{\mathcal{S}} \leqslant 1) \\
&\leqslant \mathbb{E}[\textstyle\sum_{j=1}^{r} \|\mu_{j-1} - \mu_j\|_{\mathcal{S}} \mid \bar{\mathcal{D}}] + \mathbb{P}[\mathcal{D}] \qquad \text{(triangle inequality)} \\
&\leqslant \textstyle\sum_{j=1}^{N} \mathbb{E}[\|\mu_{j-1} - \mu_j\|_{\mathcal{S}} \mid \bar{\mathcal{D}}] + \mathbb{P}[\mathcal{D}]. \qquad (7.6)
\end{aligned}$$

https://doi.org/10.1017/S0963548320000255 Published online by Cambridge University Press

For $e_j = \{v_j, u_j\}$, we let

$$\mathcal{A}_j = \max_{\sigma,\tau\in[k]^{v_j}} \|\mu_j(\cdot \mid \sigma) - \mu_j(\cdot \mid \tau)\|_{\mathcal{S}\cup\{u_j\}},$$

$$\mathcal{B}_j = \max_{\sigma,\tau\in[k]^{v_j}} \|\mu_j(\cdot \mid \sigma) - \mu_j(\cdot \mid \tau)\|_{\{u_j\}},$$

$$\mathcal{C}_j = 2 \max_{\sigma\in[k]^{e_j}} \{(\mu_j(\sigma(u_j) \mid \sigma(v_j)))^{-2}\}.$$

Letting $\mathcal{E}$ be the event that $\mathcal{C}_j \geqslant 4k^2$, we have

$$\begin{aligned}\mathbb{E}[\|\mu_{j-1} - \mu_j\|_{\mathcal{S}} \mid \bar{\mathcal{D}}] &\leqslant \mathbb{E}[\|\mu_{j-1} - \mu_j\|_{\mathcal{S}} \mid \bar{\mathcal{D}}, \bar{\mathcal{E}}] + \mathbb{P}[\mathcal{E} \mid \bar{\mathcal{D}}] \quad (\text{since } \|\mu_{j-1} - \mu_j\|_{\mathcal{S}} \in [0,1])\\ &\leqslant \mathbb{E}[\mathcal{C}_j(\mathcal{A}_j + \mathcal{B}_j) \mid \bar{\mathcal{D}}, \bar{\mathcal{E}}] + \mathbb{P}[\mathcal{E} \mid \bar{\mathcal{D}}] \qquad (7.7)\\ &\leqslant 4k^2\mathbb{E}[(\mathcal{A}_j + \mathcal{B}_j) \mid \bar{\mathcal{D}}, \bar{\mathcal{E}}] + \mathbb{P}[\mathcal{E} \mid \bar{\mathcal{D}}]. \qquad (7.8)\end{aligned}$$

We get (7.7) by working as in Lemma 4.3.

As far as $\mathbb{E}[(\mathcal{A}_j + \mathcal{B}_j) \mid \bar{\mathcal{D}}, \bar{\mathcal{E}}]$ is concerned, we work as follows. It is elementary to show that

$$\mathbb{E}[(\mathcal{A}_j + \mathcal{B}_j) \mid \bar{\mathcal{D}}, \bar{\mathcal{E}}] = (\mathbb{P}[\bar{\mathcal{D}}, \bar{\mathcal{E}}])^{-1}(\mathbb{E}[\mathcal{A}_j \mid \bar{\mathcal{D}}] + \mathbb{E}[\mathcal{B}_j \mid \bar{\mathcal{D}}]), \qquad (7.9)$$

where we use the linearity of expectation.

The expectation of both quantities $\mathcal{A}_j$ and $\mathcal{B}_j$, on the right-hand side of (7.9), depend on the subgraph of G_j which is outside $\texttt{Ball}_{\hbar-1}(w_1)$, *i.e.* the ball of radius $\hbar$ centred at w_1. This holds since e_j is outside this ball.

Let $\tilde{G}_j$ be the subgraph of G_j that is induced by the vertices outside $\texttt{Ball}_{\hbar-1}(w_1)$. Setting $\mathcal{J} = \mathcal{S} \cup \{u_j\}$, Theorem 5.1 implies that

$$\mathbb{E}[\mathcal{A}_j \mid \bar{\mathcal{D}}] \leqslant \mathbb{P}[\mathcal{D}_t(\tilde{G}_j, v_j, \mathcal{J}) > 0 \mid \bar{\mathcal{D}}],$$

where $t = 1$. Let $\tilde{G}$ be the subgraph of $G(n, d/n)$ that is induced by the vertices outside $\texttt{Ball}_{\hbar-1}(w_1)$. Then the above and Proposition 6.5 imply that

$$\mathbb{E}[\mathcal{A}_j \mid \bar{\mathcal{D}}] \leqslant \mathbb{P}[\mathcal{D}_t(\tilde{G}, v_j, \mathcal{J}) > 0 \mid \bar{\mathcal{D}}]. \qquad (7.10)$$

More specifically, Proposition 6.5 implies that

$$\mathbb{P}[\mathcal{D}_t(\tilde{G}_j, v_j, \mathcal{J}) > 0 \mid \bar{\mathcal{D}}] \leqslant \mathbb{P}[\mathcal{D}_t(\tilde{G}, v_j, \mathcal{J}) > 0 \mid \bar{\mathcal{D}}].$$

Furthermore, we have that $\tilde{G}$ is an instance of $G(N, d/n)$, where $N = n - |\texttt{Ball}_{\hbar-1}(w_1)|$, since revealing $\texttt{Ball}_{\hbar-1}(w_1)$ does not give any information for the rest of the graph. With this observation and working as in Lemma 6.6, we obtain

$$\mathbb{E}[\mathcal{A}_j \mid \bar{\mathcal{D}}] \leqslant n^{-(1/7+\log(1+\varepsilon/4)/(8\log d))}, \qquad (7.11)$$

where we use the fact that $|\mathcal{S}| \leqslant n^{1/\omega(1)}$ and the distance between v_i and $\mathcal{S}$ is $h = \log n/(7(\log d))$.

With very similar arguments we obtain

$$\mathbb{E}[\mathcal{B}_j \mid \mathcal{D}] \leqslant (20k^2\varepsilon^{-1}) \cdot n^{-(1+\log(1+\varepsilon/4)/(10(\log d)^2))}. \qquad (7.12)$$

Furthermore, working as in Corollary 6.7, for large d we get

$$\mathbb{P}[\mathcal{E} \mid \bar{\mathcal{D}}] \leqslant n^{-9/10}. \qquad (7.13)$$

Combining (7.13), (7.12), (7.11), (7.9) and (7.8), we obtain

$$\mathbb{E}[\|\mu - \nu\|_{\Lambda_j}] \leqslant n^{-\log(1+\varepsilon/4)/(9\log d)}.$$

https://doi.org/10.1017/S0963548320000255 Published online by Cambridge University Press

In light of the above, Markov's inequality implies that

$$\mathbb{P}[\|\mu - \nu\|_{\Lambda_j} \geqslant n^{-\log(1+\varepsilon/4)/(18\log d)}] \leqslant n^{-\log(1+\varepsilon/4)/(18\log d)}.$$

Then the lemma follows by applying a union bound over $j = 1, \ldots, |\Lambda|$ and noting that $|\Lambda| = o(\log n)$.

8. Proof of Theorem 2.3

For $G(n, d/n)$ consider $\mathcal{G}(G(n, d/n)) = G_0, \ldots, G_r$ such that the term G_0 is not empty but contains all the edges that belong to a cycle of length less than ℓ. The term G_r is identical to $G(n, d/n)$. Let $\{w_i, u_i\}$ be the pair of vertices which is not joined at G_i but becomes adjacent in G_{i+1}. The way we consider the sequence $\mathcal{G}(G(n, d/n))$ implies that for $i = 0, \ldots, r-1$, the pair $\{w_i, u_i\}$ does not belong to a cycle of length less than ℓ.

Letting μ_i be the Gibbs distribution over the k-colourings of G_i, letting $\mathcal{E}_i$ be the event that w_i, u_i are assigned a different colour and working as in Lemma 4.1, we have

$$\log Z(G(n, d/n), k) = \log Z(G_0, k) + \sum_{i=0}^{r-1} \log \mu_i(\mathcal{E}_i). \tag{8.1}$$

Towards proving the theorem we use the following two results.

Recall that for some integer $z > 0$ we let K_z be the number of cycles of length z in $G(n, d/n)$. Also, we let N be the number of edges in $G(n, d/n)$ which belong to a cycle of length less than ℓ. We will show that, with probability at least $1 - n^{-2/3}$ over the graph instances, we have

$$\log Z(G_0, k) = n \log k + N \log(1 - k^{-1}) + \mathcal{F}_{k,\ell}(G(n, d/n)). \tag{8.2}$$

When there is no danger of confusion, in what follows we abbreviate $\mathcal{F}_{k,\ell}(G(n, d/n))$ to $\mathcal{F}$.

Also, we will show that, with probability at least $1 - 24k^2 d\varepsilon^{-1}(1 + \varepsilon/4)^{-\ell/2}$, we have

$$\left| \sum_{i=0}^{r-1} \log \mu_i(\mathcal{E}_i) - r \log(1 - k^{-1}) \right| \leqslant (1 + \varepsilon/4)^{-\ell/2}. \tag{8.3}$$

The theorem follows immediately from (8.1), (8.2) and (8.3) and noting that $r + N$ is equal to the number of edges in $G(n, d/n)$. It remains to show that (8.2) and (8.3) are indeed correct.

We start with (8.2). Let $\mathcal{B}_1$ be the event that there is no pair of cycles of length less than ℓ that share vertices in $G(n, d/n)$. Note that if the event $\mathcal{B}_1$ does not hold, then there should be a set of at most 2ℓ vertices which spans at least $2\ell + 1$ edges. But since $\ell \leqslant \log n/(12 \log d)$, Claim 1 implies that it is unlikely to have such a set of vertices. In particular, Claim 1 implies that

$$\mathbb{P}[\mathcal{B}_1] \geqslant 1 - n^{-3/4}. \tag{8.4}$$

In what follows we assume that for $G(n, d/n)$ the event $\mathcal{B}_1$ holds.

Consider a cycle C of length z and let $w_1, w_2, \ldots, w_z$ be an ordering of the vertices such that any two consecutive vertices are adjacent, and w_1 and w_z are adjacent too. Let T_z signify the number of all possible k-colourings of C. T_z is equal to the number of all sequences $(t_1, \ldots, t_z)$ such that $t_i \in [k]$ and $t_{i+1} \neq t_i$ for all $1 \leqslant i < z$ and $t_z \neq t_1$. Let $T_1 = 0$. Then T_z satisfies the recurrence $T_z + T_{z-1} = k(k-1)^{z-1}$. To see this, observe that $k(k-1)^\ell$ is the number of all sequences $(t_1, \ldots, t_z)$ such that $t_{i+1} \neq t_i$ for all $1 \leqslant i < z$. Any such sequence either satisfies $t_z \neq t_1$, which is accounted for T_z, or $t_z = t_1$ and $t_{z-1} \neq t_1$, in which case it is contained in T_{z-1}. Hence

$$T_z = (k-1)^z + (-1)^z(k-1) = (k-1)^z \left(1 + \frac{(-1)^z}{(k-1)^{z-1}}\right).$$

https://doi.org/10.1017/S0963548320000255 Published online by Cambridge University Press

Since the cycles in G_0 are disjoint, the above implies that the number of ways the cycles in G_0 can be k-coloured is

$$\prod_{z=3}^{\ell-1}(T_z)^{K_z} = (k-1)^N \prod_{z=3}^{\ell-1}\left(1+\frac{(-1)^z}{(k-1)^{z-1}}\right)^{K_z}, \tag{8.5}$$

where we recall that N is the number of edges in G_0 which belong to a cycle of length less than ℓ. Furthermore, noting that the number of k colourings of the isolated vertices in G_0 is k^{n-N}, we obtain

$$\begin{aligned}\log(Z(G_0,k)) &= (n-N)\log k + N\log(k-1) + \sum_{z=3}^{\ell-1} K_z \log\left(1+\frac{(-1)^z}{(k-1)^{z-1}}\right)\\ &= n\log k + N\log(1-k^{-1}) + \mathcal{F}.\end{aligned}$$

Recall that the above is true when the event $\mathcal{B}_1$ holds for $G(n,d/n)$. From (8.4) and the above equation, it is immediate that (8.2) follows.

It remains to prove that (8.3) is true. Let $\mathcal{B}_2$ be the event that the number of edges in $G(n,d/n)$ is at most $J = dn/2 + n^{2/3}$. Also, let $\mathcal{Q}$ be the event that $G(n,d/n)$ is such that the inequality in (8.3) is not true. We have

$$\mathbb{P}[\mathcal{Q}] \leqslant \mathbb{P}[\mathcal{Q} \mid \mathcal{B}_2] + \mathbb{P}[\bar{\mathcal{B}}_2]. \tag{8.6}$$

As far as $\mathbb{P}[\bar{\mathcal{B}}_2]$ is concerned, applying Chernoff's bound we get

$$\mathbb{P}[\bar{\mathcal{B}}_2] \leqslant \exp(-n^{1/5}). \tag{8.7}$$

Our focus now is on $\mathbb{P}[\mathcal{Q} \mid \mathcal{B}_2]$. We show that the quantity on the left-hand side of (8.3) is typically small, by arguing that $\sum_i |\mu_i(\mathcal{E}_i) - (1-k^{-1})|$ is small, *i.e.* the difference without the logarithms.

Let $\mathcal{W}$ be the event that

$$\sum_{i=0}^{r-1} |\mu_i(\mathcal{E}_i) - (1-k^{-1})| \geqslant \frac{1}{3}(1+\varepsilon/4)^{-\ell/2}.$$

Then (8.3) follows by showing that

$$\mathbb{P}[\bar{\mathcal{Q}} \mid \mathcal{B}_2] \geqslant \mathbb{P}[\bar{\mathcal{W}} \mid \mathcal{B}_2] \quad \text{and} \quad \mathbb{P}[\mathcal{W} \mid \mathcal{B}_2] \leqslant 24k^2 d\varepsilon^{-1}(1+\varepsilon/4)^{-\ell/2}. \tag{8.8}$$

For $\mathbb{P}[\bar{\mathcal{Q}} \mid \mathcal{B}_2] \geqslant \mathbb{P}[\bar{\mathcal{W}} \mid \mathcal{B}_2]$, it suffices to show that, on the event $\mathcal{B}_2$, the event $\bar{\mathcal{W}}$ implies $\bar{\mathcal{Q}}$.

Working as in the proof of Proposition 4.2, *i.e.* (4.3) and (4.4), we get

$$\left|\sum_{i=0}^{r-1}\log\mu_i(\mathcal{E}_i) - r\log(1-k^{-1})\right| \leqslant \sum_{i=0}^{r-1}\left(\frac{|\mu_i(\mathcal{E}_i)-(1-k^{-1})|}{\mu_i(\mathcal{E}_i)} + \frac{|\mu_i(\mathcal{E}_i)-(1-k^{-1})|}{(1-k^{-1})}\right). \tag{8.9}$$

Furthermore, assuming that the event $\bar{\mathcal{W}}$ holds, we have

$$|\mu_i(\mathcal{E}_i) - (1-k^{-1})| \leqslant \sum_{i>0} |\mu_i(\mathcal{E}_i) - (1-k^{-1})| < \frac{1}{3}(1+\varepsilon/4)^{-\ell/2}, \tag{8.10}$$

where the last derivation uses the fact that the event $\bar{\mathcal{W}}$ holds. Combining (8.9) and (8.10), we have

$$\begin{aligned}\left|\sum_{i=0}^{r-1}\log\mu_i(\mathcal{E}_i) - r\log(1-k^{-1})\right| &\leqslant \sum_{i=0}^{r-1}\left(\frac{|\mu_i(\mathcal{E}_i)-(1-k^{-1})|}{1-k^{-1}-(1+\varepsilon/4)^{-\ell/2}} + \frac{|\mu_i(\mathcal{E}_i)-(1-k^{-1})|}{(1-k^{-1})}\right)\\ &\leqslant 3\sum_{i=0}^{r-1}|\mu_i(\mathcal{E}_i)-(1-k^{-1})|\\ &\leqslant (1+\varepsilon/4)^{-\ell/2}.\end{aligned} \tag{8.11}$$

https://doi.org/10.1017/S0963548320000255 Published online by Cambridge University Press

For the first inequality we use (8.10). In the second inequality we use the fact that $(1+\varepsilon/4)^{-\ell/2} \leqslant 1/4$, since

$$\ell \geqslant 2\varepsilon^{-2}\log k \geqslant \frac{4\log 2}{\log(1+\varepsilon/4)}.$$

In light of (8.11), it only remains to prove that the rightmost inequality in (8.8) is true. Letting X be a random k-colouring of G_i we have

$$\begin{aligned}
|\mu_i(\mathcal{E}_i) - (1-k^{-1})| &= |\mathbb{P}[X(w_i) = X(u_i)] - k^{-1}| \\
&\leqslant \max_{\sigma\in[k]} |\mathbb{P}[X(u_i)=\sigma \mid X(w_i)=\sigma] - \mathbb{P}[X(u_i)=\sigma]| \\
&\leqslant \max_{\sigma,\eta\in[k]^{\{w_i\}}} \|\mu_i(\cdot\mid\sigma) - \mu_i(\cdot\mid\eta)\|_{\{u_i\}} \\
&\leqslant \mathbb{P}[\mathcal{D}_\ell(G_i, w_i, u_i) > 0] \quad \text{(from Theorem 5.1)}, \qquad (8.12)
\end{aligned}$$

where the probability term in the last inequality is with respect to the Gibbs distribution on G_i. For the last inequality we use that the minimum distance between u_i, w_i is ℓ. We define the quantity

$$\rho_i = \mathbb{P}[\mathcal{D}_\ell(G_i, w_i, u_i) > 0].$$

We will bound $\mathbb{E}[\rho_i]$, where the expectation is with respect to the instance G_i. We work as in the proof of Lemma 6.6.

For what follows, we consider a uniformly random edge $\mathbf{e} \in G(n, d/n)$. Let $\mathbf{x}, \mathbf{y}$ be the two ends of $\mathbf{e}$. We have that $\mathbf{x}, \mathbf{y}$ are two random vertices in $G(n, d/n)$. Finally, let $\mathcal{A}$ be the event that $\mathbf{x}, \mathbf{y}$ are such that their distance is at least ℓ, while none of them belongs to a short cycle. It is elementary to verify that $\mathbb{P}[\mathcal{A}] = 1 - o(1)$.

We consider the graph $\boldsymbol{G} = G(n, d/n) \setminus \{\mathbf{e}\}$, *i.e.* $G(n, d/n)$ without the random edge $\mathbf{e}$. We have

$$\begin{aligned}
\mathbb{E}[\rho_i] &= \mathbb{E}_{\boldsymbol{G}}[\mathbb{P}[\mathcal{D}_\ell(\boldsymbol{G}, \mathbf{x}, \mathbf{y}) > 0] \mid \mathcal{A}] \qquad (8.13)\\
&\leqslant \mathbb{P}[\mathcal{A}]^{-1}\mathbb{E}_{\boldsymbol{G}}[\mathbb{P}[\mathcal{D}_\ell(\boldsymbol{G}, \mathbf{x}, \mathbf{y}) > 0]] \\
&\leqslant \mathbb{P}[\mathcal{A}]^{-1}\mathbb{E}_{\boldsymbol{G}}[\mathbb{E}_\mu[\mathcal{D}_\ell(\boldsymbol{G}, \mathbf{x}, \mathbf{y})]] \qquad (8.14)\\
&\leqslant \mathbb{P}[\mathcal{A}]^{-1}\mathbb{E}[\mathcal{D}_\ell(\boldsymbol{G}, \mathbf{x}, \mathbf{y})]]. \qquad (8.15)
\end{aligned}$$

The probability term on the right-hand side of (8.13) is with respect to the Gibbs distribution on G_i. In (8.14) $\mathbb{E}_\mu[\,\cdot\,]$ is the expectation with respect to the measure μ, for a given $\boldsymbol{G}$ and $\mathbf{x}, \mathbf{y}$. Equation (8.14) follows from Markov's inequality. Using Proposition 6.5, we have

$$\mathbb{E}[\mathcal{D}_\ell(\mathbf{G}, \mathbf{x}, \mathbf{y})] \leqslant (10k^2\varepsilon^{-1})n^{-1}(1+\varepsilon/4)^{-\ell}. \qquad (8.16)$$

Then, from (8.15) and (8.16) and the fact that $\mathbb{P}[\mathcal{A}] = 1 - o(1)$, we get

$$\mathbb{E}[\rho_i] \leqslant (12k^2\varepsilon^{-1})n^{-1}(1+\varepsilon/4)^{-\ell}. \qquad (8.17)$$

Furthermore, since we always have $\rho_i \geqslant 0$, we obtain

$$\mathbb{E}[\rho_i \mid \mathcal{B}_2] \leqslant \frac{\mathbb{E}[\rho_i]}{\mathbb{P}[\mathcal{B}_2]} \leqslant (14k^2\varepsilon^{-1})n^{-1}(1+\varepsilon/4)^{-\ell}, \qquad (8.18)$$

where the last inequality follows from (8.7). Combining the above bound with (8.12), we have

$$\mathbb{E}[|\mu_i(\mathcal{E}_i) - (1-k^{-1})| \mid \mathcal{B}_2] \leqslant (14k^2\varepsilon^{-1})n^{-1}(1+\varepsilon/4)^{-\ell}. \qquad (8.19)$$

https://doi.org/10.1017/S0963548320000255 Published online by Cambridge University Press

Employing the linearity of expectation, we have

$$\mathbb{E}\left[\sum_{i=0}^{r-1} |\mu_i(\mathcal{E}_i) - (1-k^{-1})| \mid \mathcal{B}_2\right] \leqslant (8dk^2\varepsilon^{-1})(1+\varepsilon/4)^{-\ell}. \tag{8.20}$$

The desired bound for $\mathbb{P}[\mathcal{W} \mid \mathcal{B}_2]$ in (8.8) follows by combining the relation above with Markov's inequality. The theorem follows.

Acknowledgement

The author would like to thank the anonymous reviewers as well as Professor Mike Paterson for the time they spent reading this manuscript and their suggestions for improving the presentation of the result. Also, the author would like to thank Amin Coja-Oghlan for the numerous fruitful discussions.

References

[1] Achlioptas, D. and Coja-Oghlan, A. (2008) Algorithmic barriers from phase transitions. In *49th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 793–802.
[2] Aldous, D. (1983) Random walks of finite groups and rapidly mixing Markov chains. In *Séminaire de Probabilités XVII 1981/82*, Springer, pp. 243–297.
[3] Bandyopadhyay, A. and Gamarnik, D. (2008) Counting without sampling: new algorithms for enumeration problems using statistical physics. *Random Struct. Algorithms* **33** 452–479.
[4] Bapst, V., Coja-Oghlan, A. and Efthymiou, C. (2017) Planting colourings silently. *Combin. Probab. Comput.* **26** 338–366.
[5] Bapst, V., Coja-Oghlan, A., Hetterich, S., Rassmann F. and Vilenchik, D. (2017) The condensation phase transition in random graph coloring. *Commun. Math. Phys.* **341** 543–606.
[6] Coja-Oghlan, A., Efthymiou, C. and Jaafari, N. (2018) Local convergence of random graph colorings. *Combinatorica* **38** 341–380.
[7] Efthymiou, C. (2015) Reconstruction/non-reconstruction thresholds for colourings of general Galton–Watson trees. In *Approximation, Randomization, and Combinatorial Optimization: Algorithms and Techniques (APPROX–RANDOM)*, pp. 756–774.
[8] Efthymiou, C. (2016) A simple algorithm for sampling colorings of $G(n, d/n)$ up to the Gibbs uniqueness threshold. *SIAM J. Comput.* **45** 2087–2116.
[9] Efthymiou, C., Hayes, T., Štefankovič, D. and Vigoda, E. (2018) Sampling random colorings of sparse random graphs. In *29th Annual ACM–SIAM Symposium on Discrete Algorithms (SODA '18)*, pp. 1759–1771.
[10] Janson, S. (1995) Random regular graphs: asymptotic distributions and contiguity. *Combin. Probab. Comput.* **4** 369–405.
[11] Jerrum, M. and Sinclair, A. (1986) Approximate counting, uniform generation and rapidly mixing Markov chains. *Inform. Comput.* **82** 93–133.
[12] Jerrum, M. and Sinclair, A. (1996) The Markov chain Monte Carlo method: an approach to approximate counting and integration. In *Approximation Algorithms for NP-hard problems* (D. S. Hochbaum, ed.), PWS.
[13] Jerrum, M., Valiant, L. G. and Vazirani, V. V. (1986) Random generation of combinatorial structures from a uniform distribution. *Theoret. Comput. Sci.* **43** 169–188.
[14] Krzakala, F., Montanari, A., Ricci-Tersenghi, F., Semerjianc, G. and Zdeborova, L. (2007) Gibbs states and the set of solutions of random constraint satisfaction problems. *Proc. Nat. Acad. Sci.* **104** 10318–10323.
[15] Mézard, M. and Montanari, A. (2009) *Information, Physics and Computation*, Oxford University Press.
[16] Mézard, M., Parisi, G. and Zecchina, R. (2002) Analytic and algorithmic solution of random satisfiability problems. *Science* **297** 812–815.
[17] Montanari, A. and Shah, D. (2007) Counting good truth assignments of random k-SAT formulae. In *18th Annual ACM–SIAM Symposium on Discrete Algorithms (SODA '07)*, pp. 1255–1264.
[18] Mossel, E. and Sly, A. (2009) Rapid mixing of Gibbs sampling on graphs that are sparse on average. *Random Struct. Algorithms* **35** 250–270.
[19] Restrepo, R., Shin, J., Tetali, P., Vigoda, E. and Yang, L. (2013) Improved mixing condition on the grid for counting and sampling independent sets. *Probab. Theory Related Fields* **156** 75–99.
[20] Robinson, R. and Wormald, N. (1994) Almost all regular graphs are Hamiltonian. *Random Struct. Algorithms* **5** 363–374.
[21] Sinclair, A., Srivastava, P. and Thurley, M. (2014) Approximation algorithms for two-state anti-ferromagnetic spin systems on bounded degree graphs. *J. Statist. Phys.* **155** 666–686.

https://doi.org/10.1017/S0963548320000255 Published online by Cambridge University Press

[22] Sinclair, A., Srivastava, P. and Yin, Y. (2013) Spatial mixing and approximation algorithms for graphs with bounded connective constant. In *54th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 300–309.
[23] Vigoda, E. (2000) Improved bounds for sampling colorings. *J. Math. Phys.* **41** 1555–1569.
[24] Wainright, M. J. and Jordan, M. (2008) Graphical models, exponential families and variational inference. In *Found. Trends Mach. Learn.* **1** (1–2).
[25] Weitz, D. (2006) Counting independent sets up to the tree threshold. In *38th Annual ACM Symposium of Theory of Computing (STOC '06)*, pp. 140–149.
[26] Yin, Y. and Zhang, C. (2016) Sampling in Potts model on sparse random graphs. In *Approximation, Randomization, and Combinatorial Optimization: Algorithms and Techniques (APPROX–RANDOM)*, pp. 1–22.

Cite this article: Efthymiou C (2020). Deterministic counting of graph colourings using sequences of subgraphs. *Combinatorics, Probability and Computing* **29**, 555–586. https://doi.org/10.1017/S0963548320000255

https://doi.org/10.1017/S0963548320000255 Published online by Cambridge University Press