On Sampling Symmetric Gibbs Distributions on Sparse Random Graphs and Hypergraphs

Charilaos Efthymiou ✉
Computer Science, University of Warwick, Coventry, UK

Abstract

In this paper, we present a novel, polynomial time, algorithm for approximate sampling from symmetric Gibbs distributions on the sparse random graph and hypergraph. The examples of symmetric distributions we consider here include some important distributions on spin-systems and spin-glasses. These are: the q-state antiferromagnetic Potts model for $q \geq 2$, including the (hyper)graph Ising model and random colourings. The uniform distribution over the Not-All-Equal solutions of a random k-SAT formula. Finally, we consider sampling from the *spin-glass* distribution called the k-spin model, i.e., this is the "diluted" version of the well-known Sherrington-Kirkpatrick model. Spin-glasses give rise to very intricate distributions which are also studied in mathematics, in neural computation, computational biology and many other areas. To our knowledge, this is the first rigorously analysed efficient sampling algorithm for spin-glasses which operates in a non trivial range of the parameters of the distribution.

We present, what we believe to be, an elegant sampling algorithm. Our algorithm is unique in its approach and does not belong to any of the well-known families of sampling algorithms. We derive it by investigating the power and the limits of the approach that was introduced in [Efthymiou: SODA 2012] and combine it, in a novel way, with powerful notions from the Cavity method.

Specifically, for a symmetric Gibbs distribution μ on the random (hyper)graph whose parameters are within an appropriate range, our sampling algorithm has the following properties: with probability $1-o(1)$ over the instances of the input (hyper)graph, it generates a configuration which is distributed within total variation distance $n^{-\Omega(1)}$ from μ. The time complexity is $O((n \log n)^2)$, where n is the size of the input (hyper)graph.

We make a notable progress regarding impressive predictions of physicists relating phase-transitions of Gibbs distributions with the efficiency of the corresponding sampling algorithms. For most (if not all) the cases we consider here, our algorithm outperforms by far any other sampling algorithms in terms of the permitted range of the parameters of the Gibbs distributions.

The use of notions and ideas from the Cavity method provides a new insight to the sampling problem. Our results imply that there is a lot of potential for further exploiting the Cavity method for algorithmic design.

2012 ACM Subject Classification Theory of computation → Design and analysis of algorithms; Theory of computation → Randomness, geometry and discrete structures; Mathematics of computing → Discrete mathematics

Keywords and phrases spin-system, spin-glass, sparse random (hyper)graph, approximate sampling, efficient algorithm

Digital Object Identifier 10.4230/LIPIcs.ICALP.2022.57

Category Track A: Algorithms, Complexity and Games

Related Version *Full Version*: https://arxiv.org/abs/2007.07145

Funding *Charilaos Efthymiou*: EPSRC New Investigator Award (grant no. EP/V050842/1) and Centre of Discrete Mathematics and Applications (DIMAP), The University of Warwick.

Acknowledgements The author of this work would like to thank Amin Coja-Oghlan for the fruitful discussions.

49th International Colloquium on Automata, Languages, and Programming (ICALP 2022).
Editors: Mikołaj Bojańczyk, Emanuela Merelli, and David P. Woodruff;
Article No. 57; pp. 57:1–57:16

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

Random constraint satisfaction problems (r-CSPs) have been the subject of intense study in combinatorics, computer science and statistical physics. In computer science the study of random CSPs is motivated by a wealth of applications, e.g., they are used as algorithmic benchmarks for hard problems such as the graph colouring, or the k-SAT, they are studied as models for statistical inference, they are also used as gadgets for cryptographic constructions, or reductions in complexity theory [23, 20, 19, 26].

Physicists, independently, have been studying random CSPs as models of *disordered systems* using the so-called *Cavity Method* (e.g. see [32, 29]). The Cavity method originates from the groundbreaking ideas in physics which got Giorgio Parisi the 2021 Nobel Prize in Physics. With its very impressive predictions and its deep intuition, the Cavity Method attracted the interest of both computer scientists and mathematicians, despite its lack of mathematical rigour. In the last two-decades, or so, ideas from the Cavity method have blended the study of random CSPs in computer science and have yield some beautiful results and breakthroughs in the area e.g., [11, 1, 34, 12, 24].

A fundamental notion in physicists' predictions is that of the Gibbs distribution. Using the Cavity method, physicists make predictions relating phase-transitions of Gibbs distributions with the efficiency of the sampling algorithms. Establishing rigorously these connections is a very challenging task and, despite any recent advances, most of the central questions still remain open. In this work, we introduce a novel approach to the sampling problem that exploits *intuition* from the Cavity method, as well as *mathematical tools* and ideas that were developed for the study of random CSPs in conjunction with the Cavity method. Our approach yields efficient algorithms with notable performance with respect to the allowed region of the parameters of the problem.

We remark that this is not the first attempt to employ ideas from the Cavity method for algorithmic design. The celebrated heuristics *Belief Propagation* and *Survey Propagation* [5, 33] are prominent examples of physicists' attempt to turn the Cavity method into an algorithm. Despite their impressive empirical performance, we lack a rigorous mathematical analysis for these heuristics with respect to generating samples from Gibbs distributions.

Our algorithm here is for sampling from what we call *symmetric Gibbs distributions*. This includes important examples such as the (hyper)graph q-colourings and its generalisation the q-state Potts model for $q \geq 2$, the symmetric variants of k-SAT such as the not-all-equal k-SAT (k-NAE-SAT). A notable case is the *spin-glass* k-spin model, i.e., the same spin-glasses that Parisi studied in the 80's. Spin-glasses give rise to very intricate distributions which are also studied in mathematics, in neural computation, computational biology and many other areas [37]. For us, the underlying geometry is an instance of the random graph, or hypergraph of constant expected degree $d > 0$.

For most of the above distributions it is extremely challenging to sample from. This is not only because the underlying geometry is random. Each of the aforementioned distributions exhibits special features that make the analysis of known sampling techniques intractable. E.g., in the interesting region of parameters for k-NAE SAT, or hypergraph colourings, we have untypical configurations with *non-local* freezing of the variables, the spin-glasses exhibit local, randomly induced, *frustration* phenomena, etc.

An additional motivation for this work comes from our desire to investigate the power and the limits of the (well-known) sampling method that is introduced in [15]. The method in [15], (or any of its subsequent works) does not exploit the Cavity method. On a high level, the approach summarises as follows: having the graph G at the input, the algorithm initially

removes all the edges and generates a configuration for the empty graph. Then, iteratively, the algorithm puts the edges back one by one. If G_i is the subgraph we have at iteration i, our objective is to have a configuration σ_i which is distributed very close to the Gibbs distribution on G_i. The idea is to generate σ_i by *updating* appropriately the configuration of G_{i-1}, i.e., update σ_{i-1} to generate efficiently σ_i. Once all the edges are put back, the algorithm outputs the configuration of G.

The algorithm in [15], relies heavily on properties that are special to graph colourings, for this reason it is restricted to this distribution. The same holds for similar approaches in the area, i.e., the algorithm is specific to the distribution it is sampling from. This is a drawback because every time we consider a new distribution we have to design a new algorithm from scratch. With our approach here, we avoid this. We introduce a sampling algorithm that the Gibbs distribution we are sampling from is a *parameter*, e.g., similarly to Glauber dynamics.

Previous approaches in the area rely on the *correlation decay* condition called *tree-uniquess* to establish the accuracy of the algorithm. For our purposes, requiring such a condition can be too restrictive. On one hand, for many of the distributions we consider here, we are far from establishing their tree uniqueness region. Actually, for many of them it is notoriously difficult to establish it even for a very limited rage of their parameters. On the other hand, it seems that Gibbs uniqueness is too restrictive a condition for distributions on the hypergraph. With our approach here we give a new insight to the problem by showing that we can exploit notions about the Gibbs distributions that we typically encounter in the study of the Cavity method and random CSPs. For example, we use notions like the *broadcasting probabilities* encountered in the study of the extremality of Gibbs distributions for random CSPs [7, 29, 16], or the *contiguity* between the Gibbs distribution and its corresponding *teacher-student model* for the study of the so-called free energy and its fluctuations [1, 11, 30, 8].

To appreciate our results, we remark that for most of the distributions we consider here, in order to prove that, e.g., the MCMC sampler works anywhere near the region of the parameters that our algorithm allows, would require major breakthroughs in the area of Markov chains, with repercussions that go beyond the problems we consider in this work.

The reader should not confuse the bounds we get here with those for "worst-case" instances. For worst case instances, usually, the parametrisation is w.r.t. the *maximum degree* of the underlying (hyper)graph, whereas for the random (hyper)graph, the natural parametrisation is w.r.t. the *expected degree*. Typically for the random (hyper)graphs here the maximum degree is *unbounded*, i.e., $\Theta(\log n/\log\log n)$, while the expected degree d is a *fixed* number.

It is important to mention that having an algorithm which allows a lot of freedom for the parameters comes at a certain cost. The approximation guarantees of our algorithm are not equally strong to, e,g., those of the MCMC approach. That is, even though the state of the art of MCMC can be far more restrictive in the parameters it allows, it provides stronger approximation guarantees. Roughly speaking, our results is summarised as follows: for a symmetric Gibbs distribution μ on the random (hyper)graph which satisfies our set of conditions, we have an approximation sampling algorithm such that with probability $1-o(1)$ over the instances of the input (hyper)graph, it generates a configuration which is distributed within total variation distance $n^{-\Omega(1)}$ from μ. The time complexity is $O((n\log n)^2)$.

Concluding, the idea of "iteratively adding edges and updating" turns out to be quite powerful, particularly when we combine it with notions and ideas from the Cavity method. It allows us to sample from distributions that, prior to this work, was impossible to sample. Our approach leads to, what we believe to be, an elegant sampling algorithm which deviates from [15] and follow up approaches not only on the phenomena of the Gibbs distributions that it utilises, but also on its basic description. Our work shows how powerful the notions

from the Cavity method can be, i.e., even in the context of sampling algorithms. We believe that there is a lot of potential towards the direction of using ideas from the Cavity method for the sampling problem in order to get even stronger algorithms.

Results for specific distributions appear in Section 1.2. Particularly, we show results for the anti-ferro q-state Potts model on graphs and hypergraphs, for any $q \geq 2$. This includes the "zero-temperature" case (hyper)graph colourings, as well as the anti-ferro Ising model. Also, we show results for the k-NAE-SAT, for $k \geq 2$, and the *spin-glass* called the k-spin model. This is the diluted version of the celebrated and extensively studied *Sherrington-Kirkpatrick* mean field model [31, 38]. Spin-glasses are studied in mathematics, in neural computation, computational biology and many other areas [37]. To our knowledge, this is the first rigorous analysis of efficient algorithm for spin-glasses. Our algorithm is by no means restricted to the cases presented here, i.e., it applies to *any* symmetric Gibbs distribution. We choose to present the specific ones because there is a common analysis framework that we can use.

1.1 General Overview

Let the fixed k-uniform hypergraph $H_k = (V, E)$. Clearly, the graph case corresponds to having $k = 2$. A Gibbs distribution on H_k is specified by the set of spins $\mathcal{A}$ and the weight functions $(\psi_e)_{e\in E}$ such that $\psi_e : \mathcal{A}^k \to \mathbb{R}_{\geq 0}$. The Gibbs distribution $\mu = \mu_H$ is on the set of configurations $\mathcal{A}^V$, i.e., the assignments of spins to the vertices of H_k, such that each $\sigma \in \mathcal{A}^V$ gets probability measure

$$\mu(\sigma) \propto \prod_{e\in E} \psi_e(\sigma(x_{e,1}), \sigma(x_{e,2}), \ldots, \sigma(x_{e,k})),$$

where $x_{e,i}$ is the i-th vertex in the hyperedge e. We assume a predefined order for the vertices in each hyperedge. The symbol $\propto$ stands for "proportional to".

In many situations, we allow ψ_e to vary with e. For example, in k-NAE-SAT, or the k-spin model each ψ_e is chosen independently, according to a predefined distribution. For this early exposition of the results the reader may very well assume that all ψ_e's are the same and fixed.

Roughly speaking, μ is symmetric, if for any $\sigma, \tau \in \mathcal{A}^V$ such that σ can be obtained from τ by repermuting the spin classes, we have that $\mu(\sigma) = \mu(\tau)$. For example, suppose that $\mathcal{A} = \{\pm 1\}$. If μ is symmetric, then we have $\mu(\sigma) = \mu(\tau)$ for any two $\sigma, \tau \in \mathcal{A}^V$, such that $\sigma(x) = -\tau(x)$ for all $x \in V$.

The underlying (hyper)graph structure is random. We let $\boldsymbol{H} = \boldsymbol{H}(n, m, k)$ be the random k-uniform hypergraph on n vertices and m hyperedges. For the graph case, i.e., $k = 2$, we write $\boldsymbol{G}(n, m)$. The expected degree is denoted by d. We take d to be a constant, i.e., $m = \Theta(n)$. Our results hold for any $d > 0$, i.e., we *do not* require that "d is sufficiently large".

Consider a typical instance of $\boldsymbol{H}$, of expected degree d, and $\mu = \mu_{\boldsymbol{H}}$ a symmetric Gibbs distribution on $\boldsymbol{H}$. Suppose that we want to sample from μ. In what follows, we describe the key features of the algorithm we propose for sampling from μ.

The main problem we need to deal with is the update-rule, i.e., how to design a method that gives us the configuration σ_i from σ_{i-1}, while, at the same time, it is generic enough to apply to all symmetric Gibbs distributions. We introduce an approach that relies on an abstract process that is called *broadcasting*. This is a natural process, that can be used to generate samples of symmetric Gibbs distributions on trees. The update starts by changing the configuration of the vertices that are disconnected and become connected when we add the (hyper) edge in the iteration i. W.l.o.g. assume that we change the assignment of just one of these vertices. Starting from this disagreeing vertex, we employ a process that is a

reminiscent of a *coupling* between two broadcasting processes. It is possible that in this process the initial disagreements propagates to some of its neighbours and, in turn, the disagreement of the neighbours propagates to neighbours further away and so on. The update stops when no more disagreements are generated.

It is crucial for the accuracy of the algorithm that the disagreements do not propagate too much further from the initial one. We can bound the rate that the disagreements spread during the update in terms of certain quantities related to the Gibbs distribution called *broadcasting probabilities*. Particularly, we have two desiderata: (a) the broadcasting probabilities are sufficiently close with each other (b) we need to show that the configuration of the vertices that the update rule encounters, somehow, looks like the result of a broadcasting process.

For the first desideratum we just need to tune appropriately the parameters of the Gibbs distribution. As far as the second one is concerned, even though we expect it to be true in our setting, it is very difficult to establish it rigorously. To this end, we employ the so-called planting-trick. This is a technique that allows us to circumvent the problem of accessing properties of the Gibbs distribution by using a very accurate approximation of it, which is called the *Teacher-Student* model. Working with the later distribution is technically much easier and it allows to get our second desideratum.

Let us be more specific. In order to have an accurate algorithm we introduce a set of technical conditions for the Gibbs distribution μ which we call `SET`. For describing `SET`, we need to visit few basic concepts.

Using the weight functions $(\psi_e)_{e\in E}$ we considered before, we introduce the following distributions: for each $e \in E$, let $\mathfrak{m}_e$ be the distribution on $\mathcal{A}^e$, i.e., configurations on the vertices in e, such that

$$\mathfrak{m}_e(\sigma) \propto \psi_e(\sigma(x_{e,1}), \sigma(x_{e,2}), \ldots, \sigma(x_{e,k})) \qquad \forall \sigma \in \mathcal{A}^e. \tag{1}$$

From Cavity's perspective, the distributions $\{\mathfrak{m}_e\}_{e\in E}$ can be viewed as *fixed-points* of the so-called *BP equations*. The distributions $\mathfrak{m}_e$ are natural objects in our setting. Particularly we focus on the so-called *broadcasting probability* $\mathfrak{m}_e^i$, which is the distribution $\mathfrak{m}_e$ conditional on the configuration at $x_{e,1}$ being $i \in \mathcal{A}$.

Furthermore, we need to introduce the notion of *total variation distance* between distributions. Particularly, for any two distributions $\hat{\nu}$ and ν on $\mathcal{A}^V$ we let

$$||\hat{\nu} - \nu||_{tv} = (1/2) \sum_{\sigma\in\mathcal{A}^V} |\hat{\nu}(\sigma) - \nu(\sigma)|.$$

Also, we let $||\hat{\nu} - \nu||_\Lambda$ be the total variation distance of the marginals of $\hat{\nu}$ and ν on the vertex set $\Lambda \subseteq V$.

`SET` consists of two conditions which we call **B.1** and **B.2**. The condition **B.1** requires that for any two $\mathfrak{m}_e^i$, $\mathfrak{m}_e^j$, i.e, any two broadcasting probabilities of μ, their total variation distance is not too large. We say that **B.1** is satisfied with slack $\delta > 0$ if we have that

$$\max_{i,j\in\mathcal{A}} ||\mathfrak{m}_e^i - \mathfrak{m}_e^j||_\Lambda \leq \frac{1-\delta}{d(k-1)}, \qquad \text{where } \Lambda = \{x_{e,2}, x_{e,3}, \ldots, x_{e,k}\}.$$

Recall that k is the size of the hyper-edge in $\boldsymbol{H}$, while d is the expected degree. The above implies that for any two broadcasting, their total variation distance should be smaller than $(1-\delta)$ over the expected number of neighbours of a vertex in $\boldsymbol{H}$.

The condition **B.2** requires mutual *contiguity* between the Gibbs distribution μ and the so-called teacher-student model. We generate the pair $(\boldsymbol{H}^*, \boldsymbol{\sigma}^*)$ according to the teacher-student model by working as follows: choose $\boldsymbol{\sigma}^*$ randomly from $\mathcal{A}^V$. Given $\boldsymbol{\sigma}^*$, generate the

weighted random hypergraph $\boldsymbol{H}^*$ on n vertices and m edges, where the weight depends on $\boldsymbol{\sigma}^*$ and μ. Contiguity implies that the *typical properties* of the pair $(\boldsymbol{H}^*, \boldsymbol{\sigma}^*)$ are the same as those of the pair $(\boldsymbol{H}, \boldsymbol{\sigma})$, where $\boldsymbol{H} = \boldsymbol{H}(n, m, k)$ and $\boldsymbol{\sigma}$ is distributed as in μ. More formally, contiguity implies that for any sequence of events $(\mathcal{S}_n)_n$ we have that

$$\Pr[(\boldsymbol{H}, \boldsymbol{\sigma}) \in \mathcal{S}_n] = o(1) \qquad \text{iff} \qquad \Pr[(\boldsymbol{H}^*, \boldsymbol{\sigma}^*) \in \mathcal{S}_n] = o(1).$$

For the planting-trick we combine **B.1** with **B.2**. Roughly speaking, we use **B.1** to prove that the update rule is local when it is applied to a configuration that is from the teacher-student model. Then, **B.2**, i.e., contiguity, implies that the same is true for our original configuration σ_{i-1}. The reader can find further details and discussions about this in Section 2.

The region of the parameters that our algorithm is accurate are specified by SET. Employing technical arguments from [11, 8, 10], we show that, for symmetric distributions, the condition **B.1** is stronger than **B.2**. In that respect, the specifications of SET reflect exactly the restrictions that **B.1** imposes to the Gibbs distribution.

The general result we prove in this work is stated in the following two theorems. Also, see Section 1.2 for applications of these theorems on specific distributions.

▶ **Theorem 1.** *For $\delta \in (0, 1]$, for integer $k \geq 2$, for any $d \geq 1/(k-1)$ and integer $m = dn/k$ the following is true for our algorithm: Consider the random k-uniform hypergraph $\boldsymbol{H} = \boldsymbol{H}(n, m, k)$. Let $\mu = \mu_{\boldsymbol{H}}$ be a symmetric Gibbs distribution on $\boldsymbol{H}$ which satisfies* SET *with slack δ.*

With probability $1 - o(1)$, over the input instances $\boldsymbol{H}$ and weight functions on the edges of $\boldsymbol{H}$, our algorithm generates a configuration whose distribution $\bar{\mu}$ is such that

$$||\bar{\mu} - \mu||_{tv} \leq n^{-\frac{\delta}{55 \log(dk)}}.$$

As mentioned above, the theorem does not require d to be a "sufficiently large constant". We chose $d \geq 1/(k-1)$, because otherwise the underlying graph structure is very simple and the problem is trivial.

Let us remark that, even though the output error for the algorithm is polynomial small, we didn't optimise the constants at the exponent of the polynomial.

▶ **Theorem 2.** *For $k \geq 2$ and $d \geq 1/(k-1)$ and integer $m = dn/k$, consider the random k-uniform hypergraph $\boldsymbol{H} = \boldsymbol{H}(n, m, k)$. The time complexity of our algorithm on input $\boldsymbol{H}$ is $O((n \log n)^2)$.*

SET Versus Gibbs Uniqueness

As noted earlier, for almost all the cases of distributions we consider here, our algorithm outperforms by far the corresponding MCMC. However, it is natural to further characterise the region of parameters that our algorithm allows, not in terms of the performance of other algorithms, but in terms of *spatial mixing* conditions on the the k-uniform random hyper-tree where each non-leaf vertex has $(k-1) \times$ Poisson(d) children and $k \geq 2$. If not anything else, this could give further insight on the approach.

Establishing the so-called tree-uniqueness for many of the Gibbs distributions here turns out to be a notoriously difficult problem even for the regular tree, not to mention the *random* tree, or the *random* hyper-tree. Deriving such results is of independent interest and goes far beyond the scope of this work. The lack of rigorous result for uniqueness allows only for a discussion on the basis of conjectures coming (mainly) from physics.

For the graph case, it is natural to compare SET with the Gibbs tree-uniqueness condition on the Galton-Watson tree with offspring distribution the Poisson(d). At least for the antiferromagnetic Ising and Potts model and the 2-NAE-SAT model, the requirement of our algorithm coincides with the conjectured tree uniqueness region of the the random tree. We are not aware of any conjectures about the tree uniqueness of the spin-glass 2-spin model.

For the hyper-graph case, things seem to be more interesting. The author of this work is not aware of any physics' conjectures about the Gibbs uniqueness on the random hyper-tree. However, we believe that is interesting to include in our discussion the following (somehow easy to make) observation: For the sake of our discussion, rather than random hyper-tree consider a regular one, e.g. consider k-uniform hyper-tree where each non-leaf vertex has $\Delta(k-1)$ children, for integers $\Delta, k > 0$. Consider also a symmetric Gibbs distribution with hard constraints on this hyper-tree, e.g. the q-colouring model. Provided that Δ, k are relatively large, say $\Delta, k \geq 15$, the condition SET does not preclude configurations at the vertices at level ℓ of the tree that specify uniquely the colouring at the root, for howsoever large $\ell > 0$ we choose [1]. Of course, such colourings are *extremely untypical* with respect to the Gibbs distribution on the hyper-tree. A very similar phenomenon can be observed at the random hyper-tree, too. This aforementioned phenomenon leads us to *conjecture* for the hypergraph cases, that our algorithm allows to sample beyond the corresponding hyper-tree uniqueness region for *all* the Gibbs distributions we consider here.

The above should not be a surprise to the reader. It is well-known that for hard-constraint distributions on the hyper-graph (not necessarily random), the worst case configurations can be very problematic for the analysis of MCMC sampler, i.e., even under very mild conditions, e.g. see [22]. Our algorithm does not suffer from such problems, because it deals with typical configurations which are much nicer.

Related work

The idea of "iteratively adding edges and updating" for sampling was first introduced in [15] for sampling colourings of random graphs. The techniques and tools we introduce here for the sampling problem, rely on results developed in the study Cavity method and random CSP, e.g. see [1, 8, 10, 11].

There are two other works which follow the approach of "adding edges and updating" and use the same correlation decay approach to [15]. One is [17], an improvement of [15], which is about colourings of the random graph of sufficiently large expected degree d. The other one is [4] for the Potts model on the related random Δ-regular graph, for large Δ. From the second paper, it is conceivable that we can get an efficient algorithm only for the ferromagnetic Potts model on the random graph, provided that the expected degree d is large. Apart from colourings and ferro-Potts on the graph, we cannot rely on any of these two approaches for our endeavours. Both of them rely on the special properties of the distribution they are sampling from, thus they don't allow for other distributions. Furthermore, their tree-uniqueness requirement essentially restrict their use to considering only graphs, rather than hypergraphs. Our work here improves on both results in [17, 4] as it allows for any expected degree $d > 0$, i.e., rather than sufficiently large d.

[1] E.g., in this setting SET allows $q \approx ((k-1)\Delta)^{1/(k-1)}$, while the number of children of a non-leaf vertex is $(k-1)\Delta \gg q$.

There are other approaches to sampling from Gibbs distributions (not-necessarily on random graphs), which is different than the one we consider here. Notably, the most popular ones rely on the Markov Chain Monte Carlo Method (MCMC) [28, 21]. The literature of MCMC algorithms is vast and includes some beautiful results, just to mention a few [3, 39, 27, 25, 35, 14, 18, 6, 13].

Our results about the colourings are related to the work in [18] for MCMC sampling. Some result from [18] can be extended naturally to the Potts model, too. In that respect our approach outperforms, by far, [18] it terms of the range of the allowed parameters of the Gibbs distributions. However, we note that the MCMC algorithm achieves better approximation guarantees in the (more restricted) regions it operates.

1.2 Applications

1.2.1 The antiferromagnetic Ising Model

The Ising model on the k-uniform hypergraph $H_k = (V, E)$ is a distribution on the set of configurations $\{\pm 1\}^V$ such that each $\sigma \in \{\pm 1\}^V$ is assigned probability measure

$$\mu(\sigma) \propto \exp\left(\beta \cdot \sum_{e\in E} \prod_{x,y\in e} \mathbf{1}\{\sigma(x) = \sigma(y)\} + h \cdot \sum_{x\in V} \sigma(x)\right),$$

where $\beta \in \mathbb{R}$ is the *inverse temperature* and h is the *external field*. It is straightforward that the Ising model is symmetric only when $h = 0$. We assume $\beta < 0$, which corresponds to the *antiferromagnetic* Ising model.

For $\Delta, k > 0$ such that $\Delta > \frac{2^{k-1}-1}{k-1}$ we let the function

$$\beta_{\text{Ising}}(\Delta, k) = \log\left(\frac{\Delta(k-1)+1-2^{k-1}}{\Delta(k-1)+1}\right).$$

The uniqueness region of the antiferromagnetic Ising model on the Δ-ary tree, for $\Delta \geq 1$, is well-known. Particularly, it corresponds to temperatures β such that either $\Delta > 2$ and $\beta_{\text{Ising}}(\Delta, 2) < \beta < 0$, or $\Delta = 1$ and finite $\beta < 0$, i.e., $\beta \neq -\infty$.

▶ **Theorem 3.** *For integer $k \geq 2$ fr any $d \geq 1/(k-1)$ such that either*
1. *$d > (2^{k-1}-1)/(k-1)$ and $\beta_{\text{Ising}}(d, k) < \beta < 0$, or*
2. *$d \leq (2^{k-1}-1)/(k-1)$ and finite $\beta < 0$,*

the following is true: For the random k-uniform hypergraph $\boldsymbol{H} = \boldsymbol{H}(n, m, k)$, where $m = dn/k$, let $\mu = \mu_{\boldsymbol{H}}$ be the antiferromagnetic Ising model on $\boldsymbol{H}$, with inverse temperature β and external field $h = 0$.

There exists $c_0 > 0$ which depends only on the choice of k, d, β, such that with probability $1 - o(1)$ over the input instances $\boldsymbol{H}$, our algorithm generates a configuration with distribution $\bar{\mu}$ such that

$$||\bar{\mu} - \mu||_{tv} \leq n^{-\frac{c_0}{55\log(dk)}}.$$

The time complexity of the algorithm is $O((n \log n)^2)$ with probability 1.

▶ Remark 4. For the graph cases, i.e., $k = 2$, physics' conjecture is that we get the Gibbs uniqueness region for the random trees with expected offspring d, by somehow "pretending" that we are dealing with a d-ary tree (regardless of d being an integer, or not). From the above theorem, it is evident that the region that our algorithm operates, corresponds exactly to the region that is implied by the physics' non rigorous consideration. Someone could observe the same for the related Potts model in the following section.

1.2.2 The antiferromagnetic Potts Model and the Colouring Model

The Potts model on the k uniform hypergraph $H_k = (V, E)$ is a generalisation of the Ising model in the sense that it allows for q spins where $q \geq 2$. Particularly, each $\sigma \in [q]^V$, where $[q] = \{1, 2, \ldots, q\}$, is assigned probability measure

$$\mu(\sigma) \propto \exp\left(\beta \cdot \sum_{e \in E} \prod_{x,y \in e} \mathbf{1}\{\sigma(x) = \sigma(y)\}\right).$$

where $\beta \in \mathbb{R}$ is the *inverse temperature*. The graph case, i.e., $k = 2$, follows immediately from the above.

There is a version of the Potts model with external field, similarly to the Ising model. We do not consider the case with field because it gives rise to a non symmetric distribution. The antiferromagnetic Potts model we focus here corresponds to having $\beta < 0$.

A very interesting case of the Potts model is the *colouring model*. This is the uniform distribution over the *proper* q-colourings of the underlying (hyper)graph H_k, i.e., we do not allow configurations with monochromatic edges. The colouring model corresponds to the Potts model with $\beta = -\infty$.

For $\Delta, k > 0$ such that $\Delta > \frac{q^{k-1}-1}{k-1}$, we let the function

$$\beta_{\text{Potts}}(\Delta, q, k) = \log\left(\frac{\Delta(k-1)+1-q^{k-1}}{\Delta(k-1)+1}\right).$$

▶ **Theorem 5.** *For integer $k \geq 2$, for any $d \geq 1/(k-1)$, integers $q > 2$ and $m = dn/k$ the following is true: Assume that β, q satisfy one of the following cases:*

1. *$(q^{k-1} - 1)/(k-1) < d$ and $\beta_{\text{Potts}}(d, q, k) < \beta < 0$,*
2. *$(q^{k-1} - 1)/(k-1) > d$ and $\beta < 0$, including $\beta = -\infty$.*

Consider the random k-uniform hypergraph $\boldsymbol{H} = \boldsymbol{H}(n, m, k)$. Let $\mu = \mu_{\boldsymbol{G}}$ be the q-state antiferromagnetic Potts model on $\boldsymbol{H}$ with inverse temperature β. There exists $c_0 > 0$, which depends only on our choices of k, d, β such that with probability $1 - o(1)$ over the input instances $\boldsymbol{H}$, our algorithm generates a configuration whose distribution $\bar{\mu}$ is such that

$$||\bar{\mu} - \mu||_{tv} \leq n^{-\frac{c_0}{55\log(dk)}}.$$

The time complexity of the algorithm is $O((n \log n)^2)$ with probability 1.

Physics' (folklore) conjecture for the uniqueness of antiferromagnetic Potts model on the Galton Watson tree with offspring distribution Poisson(d) coincides with the region specified in the theorem above, for $k = 2$.

1.2.3 The k-NAE-SAT Model

For integer $k \geq 2$, let $\mathbf{F}_k(n, m)$ be a random propositional formula over the Boolean variables $x_1, \ldots, x_n$. Particularly, $\mathbf{F}_k(n, m)$ is obtained by inserting m independent random clauses of length k such that no variable appears twice in the same clause. Here, we consider formulas with $m = dn/k$ clauses for a fixed number d, i.e., on average every variable occurs in d clauses.

We focus on the "Not-All-Equal" satisfying assignments of $\mathbf{F}_k(n, m)$. A Boolean assignment σ of $x_1, \ldots, x_n$ is NAE-satisfying for $\mathbf{F}_k(n, m)$ if under both σ and its binary inverse $\bar{\sigma}$ all m clauses evaluate to "true". The random k-NAE-SAT problem is one of the standard examples of random CSPs and has received a great deal of attention. In particular, in an influential paper Achlioptas and Moore [2] pioneered the use of the second moment method for estimating the partition functions of random CSPs with the example of random k-NAE-SAT. Our focus is on sampling from the *uniform distribution* over the NAE satisyfing assignments of $\mathbf{F}_k(n, m)$. We have the following result.

▶ **Theorem 6.** *For $\delta \in (0,1]$, for $k \geq 2$, for any $1/(k-1) \leq d < (1-\delta)\frac{2^{k-1}-1}{k-1}$ and for integer $m = dn/k$, the following is true for our algorithm: Consider $\mathbf{F}_k(n,m)$ and let μ be the uniform distribution over the NAE satisfying assignments of $\mathbf{F}_k(n,m)$. With probability $1-o(1)$ over the input instances $\mathbf{F}_k(n,m)$, our algorithm generates a configuration whose distribution $\bar{\mu}$ is such that*

$$||\bar{\mu} - \mu||_{tv} \leq n^{-\frac{\delta}{55\log(dk)}}.$$

The time complexity of the algorithm is $O((n\log n)^2)$ with probability 1.

As a point of reference for the performance of our *sampling* algorithm, note that it works in a region of parameters which is very close to those of *search* algorithms for the problem, e.g. see the analysis for the renown *walk-sat* algorithm in [9].

1.2.4 The *k*-spin model

For integer $k \geq 2$, consider the k-uniform hypergraph $H_k = (V,E)$. Additionally, let $\mathbf{J} = (\mathbf{J}_e)_{e\in E}$ be a family of independent, standard Gaussians (expectation zero and variance one). The *k-spin model* on H_k at inverse temperature $\beta > 0$ is the distribution that assigns each configuration $\sigma \in \{\pm 1\}^V$ the probability measure

$$\mu(\sigma) \propto \prod_{\alpha\in E} \exp\left(\beta \mathbf{J}_e \prod_{y\in\alpha}\sigma(y)\right). \tag{2}$$

It is elementary to verify that the k-spin model is symmetric when $k \geq 2$ is an even integer. Here we consider the above distribution when the underlying (hyper)graph is an instance of $\boldsymbol{H} = \boldsymbol{H}(n,m,k)$ of expected degree d, i.e., $m = dn/k$.

The k-spin model is the diluted version of the well-known Sherrington-Kirkpatrick model which has been subject of intense study both in mathematics and physics [38, 31]. Generally, spin-glasses give rise to very intricate distributions and they are also studied in neural networks, computational biology and other areas of computer science [37]. As mentioned before, this is the first efficient algorithm for sampling from spin-glasses in a non-trivial region of their parameters. For what follows we consider the function

$$F_k(x) = \frac{|e^x - e^{-x}|}{(2^{k-1}-1)e^{-x} + e^x}. \tag{3}$$

▶ **Theorem 7.** *For $\delta \in (0,1]$, for even integer $k \geq 2$, for any $d \geq 1/(k-1)$ and for any $\beta \geq 0$ such that*

$$\mathbb{E}[F_k(\beta \mathbf{J}_0)] \leq \tfrac{1-\delta}{d(k-1)},$$

where the expectation is w.r.t. the standard Gaussian random variable $\mathbf{J}_0$, the following is true for our algorithm: Consider $\boldsymbol{H} = \boldsymbol{H}(n,m,k)$, where $m = dn/k$, and let μ be the k-spin model on $\boldsymbol{H}$ at inverse temperature β. With probability $1-o(1)$ over the input instances $\boldsymbol{H}$ and the weight functions on the edges of $\boldsymbol{H}$, our algorithm generates a configuration whose distribution $\bar{\mu}$ is such that

$$||\bar{\mu} - \mu||_{tv} \leq n^{-\frac{\delta}{55\log(dk)}}.$$

The time complexity of the algorithm is $O((n\log n)^2)$ with probability 1.

Notation

Let the hypegraph $H_k = (V, E)$ and the Gibbs distribution μ on the set of configurations $\mathcal{A}^V$. For a configuration σ, we let $\sigma(\Lambda)$ denote the configuration that σ specifies on the set of vertices Λ. We let μ_Λ denote the marginal of μ at the set Λ. For a configuration $\sigma \in \mathcal{A}^V$ we let $\mu(\cdot \mid \Lambda, \sigma)$, denote the distribution μ conditional on the configuration at Λ being $\sigma(\Lambda)$. Also, we interpret the conditional marginal $\mu_\Lambda(\cdot \mid \Lambda', \sigma)$, for $\Lambda' \subseteq V$, in the natural way.

2 Algorithmic Approach – High Level Description

To facilitate the exposition of the algorithm assume in this section that we are dealing with a graph, rather than hypergraph, while we assume that this graph is fixed.

First, we recall the algorithm: on input G, the algorithm initially removes all the edges and generates a configuration for the empty graph. Then, iteratively, we put the edges back one by one. If G_i is the subgraph we have at iteration i, our aim is to have a configuration σ_i which is distributed very close to the Gibbs distribution on G_i, for every i. We generate σ_i by updating appropriately σ_{i-1}, the configuration of G_{i-1}. Once all edges are put back, the algorithm outputs the configuration of G. One of the main challenge is to specify the *update rule*, i.e., how to generate σ_i from σ_{i-1}.

We describe the proposed rule by considering the following, simpler, problem. Consider two *high-girth*, fixed, graphs $G = (V, E)$ and $G' = (V, E')$. Assume that G and G' differ on a single edge, i.e. compared to G, the graph G' has the extra edge $e = \{u, w\}$. Let μ and μ' be the Gibbs distributions of G and G', respectively. We want to use the update rule to generate efficiently $\boldsymbol{\tau}$ a sample from μ', while we are given $\boldsymbol{\sigma}$, a sample from μ.

To facilitate our exposition, assume that we already know $\boldsymbol{\tau}(u)$ and $\boldsymbol{\tau}(w)$, and they are such that $\boldsymbol{\tau}(u) = \boldsymbol{\sigma}(u)$ and $\boldsymbol{\tau}(w) \neq \boldsymbol{\sigma}(w)$. In what follows, we focus on specifying $\boldsymbol{\tau}$ for the rest of the vertices in V.

The plan is to visit each vertex z *iteratively* and specify $\boldsymbol{\tau}(z)$. At each iteration t, we only know the configuration of $\boldsymbol{\tau}$ for the vertices in the set of vertices we have already visited, we call it $\mathcal{N}_t$. Initially we have that $\mathcal{N}_0 = \{w, u\}$. Also, let $\mathcal{D} = \{\boldsymbol{\tau}(w), \boldsymbol{\sigma}(w)\}$, i.e., $\mathcal{D}$ is the set of the spins of the initial disagreement. At iteration t we pick a vertex z which is outside $\mathcal{N}_t$ but has a neighbour $x \in \mathcal{N}_t$ which is disagreeing, i.e., $\boldsymbol{\tau}(x) \neq \boldsymbol{\sigma}(x)$. For the moment, assume that such vertex exists.

If $\boldsymbol{\sigma}(z) \notin \mathcal{D}$, then we just set $\boldsymbol{\tau}(z) = \boldsymbol{\sigma}(z)$. On the other hand, if $\boldsymbol{\sigma}(z) \in \mathcal{D}$, then we work as follows: there is a probability p_z, that depends on the configuration of $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ at $\mathcal{N}_t$, and we set

$$\boldsymbol{\tau}(z) = \begin{cases} \mathcal{D} \setminus \{\boldsymbol{\sigma}(z)\} & \text{with prob. } p_z \\ \boldsymbol{\sigma}(z) & \text{with prob. } 1 - p_z. \end{cases}$$

The first line implies that $\boldsymbol{\tau}(z)$ gets the opposite spin of $\boldsymbol{\sigma}(z)$. E.g., if $\mathcal{D} = \{\texttt{red}, \texttt{blue}\}$ and $\boldsymbol{\sigma}(z) = \texttt{red}$, then $\boldsymbol{\tau}(z) = \texttt{blue}$. Once $\boldsymbol{\tau}(z)$ is decided, set $\mathcal{N}_{t+1} = \mathcal{N}_t \cup \{z\}$ and continue with the next iteration.

It could be that in iteration t, there is no vertex z outside $\mathcal{N}_t$ which has a disagreeing neighbour inside $\mathcal{N}_t$. If this is the case, then for every z for which we have not specified $\boldsymbol{\tau}(z)$, we set $\boldsymbol{\tau}(z) = \boldsymbol{\sigma}(z)$. Once we have specified the assignment $\boldsymbol{\tau}$ for every vertex z in the graph, the update rule terminates.

The probability p_z is determined in terms of a *maximal coupling* between the marginals of μ' and μ at z, conditional on $\boldsymbol{\tau}(\mathcal{N}_t)$ and $\boldsymbol{\sigma}(\mathcal{N}_t)$. We denote these marginals as $\mu'_z(\cdot \mid \mathcal{N}_t, \boldsymbol{\tau})$ and $\mu_z(\cdot \mid \mathcal{N}_t, \boldsymbol{\sigma})$, respectively. We have

$$p_z = \max\left\{0, 1 - \frac{\mu'_z(\boldsymbol{\sigma}(z) \mid \mathcal{N}_t, \boldsymbol{\tau})}{\mu_z(\boldsymbol{\sigma}(z) \mid \mathcal{N}_t, \boldsymbol{\sigma})}\right\}.$$

One can show that the above generates a *perfect* sample from the distribution μ'. There is an issue with this approach, though. It is not clear how we can compute the probabilities p_z, efficiently. Computing p_z relies on estimating conditional marginals of Gibbs distributions. In our setting, we don't known how to estimate these marginals efficiently. To this end, we use *different* probabilities. That is, we follow the previous steps and when at the iteration t we examine a vertex z for which $\boldsymbol{\sigma}(z) \in \mathcal{D}$, we set $\boldsymbol{\tau}(z)$ such that

$$\boldsymbol{\tau}(z) = \begin{cases} \mathcal{D} \setminus \{\boldsymbol{\sigma}(z)\} & \text{with prob. } q_z \\ \boldsymbol{\sigma}(z) & \text{with prob. } 1 - q_z, \end{cases} \tag{4}$$

i.e., instead of p_z we use q_z. Recall that we choose z because it has a disagreeing neighbour $x \in \mathcal{N}_t$. Each q_z can be expressed in terms of the simpler distribution $\mathfrak{m}_\alpha$, where α is the edge between z and x. We have

$$q_z = \max\left\{0,\ 1 - \frac{\mathfrak{m}_{\alpha,z}(\boldsymbol{\sigma}(z) \mid x, \boldsymbol{\tau})}{\mathfrak{m}_{\alpha,z}(\boldsymbol{\sigma}(z) \mid x, \boldsymbol{\sigma})}\right\}. \tag{5}$$

Recall from our notation that $\mathfrak{m}_{\alpha,z}(\cdot \mid x, \boldsymbol{\tau})$ is the marginal of $\mathfrak{m}_\alpha$ on z, conditional on x being set $\boldsymbol{\tau}(x)$. Also, note from (1) that the distribution $\mathfrak{m}_\alpha$ is very simple and can be computed very fast.

A natural question at this point is what motivates the use q_z in the place of p_z. We observe that if our graphs G and G' were trees, then we would have that $q_z = p_z$. That is, for trees our update rule generates perfect samples from μ'. In some sense, our approach amounts to approximating the probabilities p_z, which are difficult to compute, with those of the tree, which we can compute very fast. In light of our assumption that our graphs G and G' are of high-girth, i.e., locally tree-like, this approximation seems quite natural.

Under certain conditions, our approach yields very good approximations of μ'. The update rule is accurate in settings where, typically, the set of vertices that change assignment does not grow "too large". To be more specific, let $\mathcal{Q}$ be the set of vertices that change configuration during the update, i.e., their configuration under $\boldsymbol{\tau}$ is different than that under $\boldsymbol{\sigma}$. Somehow, our update rule runs into trouble when $\mathcal{Q}$ induces a subgraph which contains one of the long cycles of G, or $\mathcal{Q}$ reaches u. In this case we consider that the algorithm *fails*. That is, our update rule outputs either a configuration $\boldsymbol{\tau} \in \mathcal{A}^V$, or a fail status. We establish a connection between the accuracy of the update and its failure probability, particularly we show that the smaller the failure probability the more accurate the algorithm is.

2.1 Accuracy and failure probabilities

We relate the approximation error of the update rule with failure probabilities by exploiting an interesting property of the update rule which is a reminiscent of the *detailed balance equation* from the theory of reversible Markov chains [36]. In what follows, first we describe the "detailed balance property" of the update and then we show how we use it to study the accuracy.

In the same setting as before, assume that $\boldsymbol{\sigma}(\{u, w\}) = \sigma$ and $\boldsymbol{\tau}(\{u, w\}) = \tau$, for fixed $\sigma, \tau \in \mathcal{A}^{\{u,w\}}$. The update rule can be viewed as a stochastic process that takes a configuration that agrees with σ at $\{u, w\}$ and generates either a new configuration which

agrees with τ at $\{u, w\}$, or fails. There is a natural way of defining the *reverse update* which works towards the opposite direction, i.e., takes a configuration which agrees with τ and either generates a configuration that agrees with σ at $\{u, w\}$, or fails.

For any two configuration $\kappa, \eta \in \mathcal{A}^V$, we let $\mathrm{P}_{\sigma,\tau}(\kappa, \eta)$ be the probability that on input κ the update generates η. Similarly we can define $\mathrm{P}_{\tau,\sigma}(\eta, \kappa)$ for the reverse update. The detailed balance equation relates these two probabilities, i.e., it specifies that

$$\mu(\kappa)\mathrm{P}_{\sigma,\tau}(\kappa, \eta) = \mu(\eta)\mathrm{P}_{\tau,\sigma}(\eta, \kappa).$$

Note that, in the equation above, the Gibbs distributions μ are unconditional.

We proceed by demonstrating how we use the detailed balance to get the update error. Consider the same setting as in the previous paragraphs. Let $\bar{\mu}$ be the distribution of the output of the update when the input is distributed as in $\mu(\cdot \mid \{u, w\}, \sigma)$.

We need to focus on the failure probability of both the update and the reverse update. Let $\mathtt{F}(\kappa)$ be the failure probability of the update rule on input (fixed) κ. In that respect, the failure probability is equal to $\mathbb{E}[\mathtt{F}(\boldsymbol{\kappa})]$, where the expectation is w.r.t. to $\boldsymbol{\kappa}$ which is distributed as in $\mu(\cdot \mid \{w, u\}, \sigma)$. Similarly, let $\mathtt{R}(\eta)$ be the failure probability for the reverse update on input (fixed) η. The failure probability of the reverse update is $\mathbb{E}[\mathtt{R}(\boldsymbol{\eta})]$, where $\boldsymbol{\eta}$ is distributed as in $\mu(\cdot \mid \{w, u\}, \tau)$.

Using the detail balance and an asymptotic independence result between the configuration of w and u under μ, we get the following: For any $\eta \in \mathcal{A}^V$ we have that

$$\begin{aligned}\bar{\mu}(\eta) &= \sum_{\kappa \in \mathcal{A}^V} \mu(\kappa \mid \{w, v\}, \sigma)\mathrm{P}_{\sigma,\tau}(\kappa, \eta) \approx \sum_{\kappa \in \mathcal{A}^V} \mu(\eta \mid \{w, v\}, \tau)\mathrm{P}_{\tau,\sigma}(\eta, \kappa) \\ &= \mu(\eta \mid \{w, v\}, \tau)(1 - \mathtt{R}(\eta)). \end{aligned} \tag{6}$$

The first equation is just the definition of $\bar{\mu}(\eta)$. The detailed balance with the asymptotic independence are used for the derivation with "$\approx$". The last equation follows from the observation that summing $\mathrm{P}_{\tau,\sigma}(\eta, \kappa)$ over $\kappa \in \mathcal{A}^V$ is equal to the probability that the reverse update does not fail, when the input is η.

Another observation that we use is the following one: if the update has a positive failure probability, then $\sum_\eta \bar{\mu}(\eta) < 1$. This holds because $\bar{\mu}$ gives positive measure to the failure status of the update. That is, we have that $\mathbb{E}[\mathtt{F}(\boldsymbol{\kappa})] + \sum_\eta \bar{\mu}(\eta) = 1$. Combining this equality and (6), standard derivations imply

$$||\bar{\mu} - \mu(\cdot \mid \{w, u\}, \tau)||_{tv} \approx (1/2)(\mathbb{E}[\mathtt{R}(\boldsymbol{\eta})] + \mathbb{E}[\mathtt{F}(\boldsymbol{\kappa})]).$$

Essentially the update error is equal to the average of the failure probabilities of the update and its reverse.

2.2 The failure probability on the random graph

Here we highlight the intuition behind our claim that if `SET` holds, then the failure probability is very small.

Consider a setting which is a bit simpler than that we had before. Let μ be a symmetric Gibbs distribution on $\boldsymbol{G} = \boldsymbol{G}(n, m)$. Consider the spins $c, c' \in \mathcal{A}$ such that $c \neq c'$. Let $\boldsymbol{\sigma}$ be distributed as in μ conditional on $\boldsymbol{\sigma}(u) = c'$, for some $u \in V$. We use the update process to generate a configuration $\boldsymbol{\tau}$ which is (approximately) distributed as in μ conditional on that $\boldsymbol{\tau}(u) = c$. Our focus is on the probability of failure for the update. Particularly, we argue that if `SET` holds, then the size of the set of disagreeing vertices in the update process, i.e, the vertices v such that $\boldsymbol{\sigma}(v) \neq \boldsymbol{\tau}(v)$, grows *subcritically* at each iteration.

Recall that, in the update, the disagreements start from vertex u and iteratively propagate over the graph. Assume that not too many vertices have been visited in the process. In iteration t, the process chooses the vertex z which is adjacent to the disagreeing vertex x, i.e., we already know that $\boldsymbol{\tau}(x) \neq \boldsymbol{\sigma}(x)$. The probability of disagreement propagating to z can be estimated by just using (4) and (5). The idea is to combine (4) and (5) with the randomness of $\boldsymbol{\sigma}$ and show that the probability of disagreement at z is $< 1/d$, which would imply the subcriticality for the update process.

Exploiting the randomness of $\boldsymbol{\sigma}$ means that at iteration t of the process we only have exposed the configuration of $\boldsymbol{\sigma}$ for the vertices which the process has already visited. If the process hasn't revealed the configuration of $\boldsymbol{\sigma}$ for too many vertices, then the marginal of the configuration at z should be close to $\mathfrak{m}_e(\cdot \mid x, \boldsymbol{\sigma})$, where $e = \{x, z\}$. This would imply that the disagreement probability at z is at most

$$\max_{c,c' \in \mathcal{A}} ||\mathfrak{m}_e(\cdot \mid x, c) - \mathfrak{m}_e(\cdot \mid x, c')||_z.$$

The above observation is quite interesting because the quantity above, i.e., disagreement probability, is upper bounded by using **B.1** in SET. Particularly, **B.1** implies that the above total variation distance is $\leq (1-\delta)/d$. Thus, if the above intuition is correct, then we can get the subcritical growth by exploiting the condition **B.1**. Unfortunately, with our assumptions about μ, it too difficult to argue that the marginal probability at z is very close to $\mathfrak{m}_e(\cdot \mid x, \boldsymbol{\sigma})$ in our process.

To this end, we employ the teacher-student model. We consider the pair $(\boldsymbol{G}^*, \boldsymbol{\sigma}^*)$ from the teacher-student model. We study the propagation of disagreements for the update process on the pair $(\boldsymbol{G}^*, \boldsymbol{\sigma}^*)$. There, it is simpler to argue that the distribution of z is very close to $\mathfrak{m}_e(\cdot \mid x, \boldsymbol{\sigma})$. The condition **B.1** still applies here and it implies that the growth of disagreements in $\boldsymbol{G}^*$ is subcritical. In turn, this implies that the failure probability for the specific update is very small. Subsequently, we employ contiguity, i.e., **B.2**, to argue that if the probability of failure for the case of $(\boldsymbol{G}^*, \boldsymbol{\sigma}^*)$ is very small, then the probability of failure for $(\boldsymbol{G}, \boldsymbol{\sigma})$ cannot be much larger.

References

1. Dimitris Achlioptas and Amin Coja-Oghlan. Algorithmic barriers from phase transitions. In *49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA*, pages 793–802. IEEE Computer Society, 2008. doi:10.1109/FOCS.2008.11.
2. Dimitris Achlioptas and Cristopher Moore. Random k-sat: Two moments suffice to cross a sharp threshold. *SIAM J. Comput.*, 36(3):740–762, 2006. doi:10.1137/S0097539703434231.
3. Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. Spectral independence in high-dimensional expanders and applications to the hardcore model. In *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*, pages 1319–1330. IEEE, 2020. doi:10.1109/FOCS46700.2020.00125.
4. Antonio Blanca, Andreas Galanis, Leslie Ann Goldberg, Daniel Stefankovic, Eric Vigoda, and Kuan Yang. Sampling in uniqueness from the potts and random-cluster models on random regular graphs. *SIAM J. Discret. Math.*, 34(1):742–793, 2020. doi:10.1137/18M1219722.
5. Alfredo Braunstein, Marc Mézard, and Riccardo Zecchina. Survey propagation: An algorithm for satisfiability. *Random Struct. Algorithms*, 27(2):201–226, 2005. doi:10.1002/rsa.20057.
6. Sitan Chen, Michelle Delcourt, Ankur Moitra, Guillem Perarnau, and Luke Postle. Improved bounds for randomly sampling colorings via linear programming. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, San Diego, California, USA, January 6-9, 2019*, pages 2216–2234. SIAM, 2019. doi:10.1137/1.9781611975482.134.

7 Amin Coja-Oghlan, Charilaos Efthymiou, and Nor Jaafari. Local convergence of random graph colorings. *Combinatorica*, 38(2):341–380, 2018. doi:10.1007/s00493-016-3394-x.

8 Amin Coja-Oghlan, Charilaos Efthymiou, Nor Jaafari, Mihyun Kang, and Tobias Kapetanopoulos. Charting the replica symmetric phase. *Communications in Mathematical Physics*, 359(2):603–698, 2018. doi:10.1007/s00220-018-3096-x.

9 Amin Coja-Oghlan and Alan M. Frieze. Analyzing walksat on random formulas. *SIAM J. Comput.*, 43(4):1456–1485, 2014. doi:10.1137/12090191X.

10 Amin Coja-Oghlan, Tobias Kapetanopoulos, and Noëla Müller. The replica symmetric phase of random constraint satisfaction problems. *Comb. Probab. Comput.*, 29(3):346–422, 2020. doi:10.1017/S0963548319000440.

11 Amin Coja-Oghlan, Florent Krzakala, Will Perkins, and Lenka Zdeborová. Information-theoretic thresholds from the cavity method. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 146–157. ACM, 2017. doi:10.1145/3055399.3055420.

12 Jian Ding, Allan Sly, and Nike Sun. Proof of the satisfiability conjecture for large k. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pages 59–68. ACM, 2015. doi:10.1145/2746539.2746619.

13 Martin E. Dyer and Alan M. Frieze. Randomly coloring graphs with lower bounds on girth and maximum degree. *Random Struct. Algorithms*, 23(2):167–179, 2003. doi:10.1002/rsa.10087.

14 Martin E. Dyer, Alan M. Frieze, Thomas P. Hayes, and Eric Vigoda. Randomly coloring constant degree graphs. *Random Struct. Algorithms*, 43(2):181–200, 2013. doi:10.1002/rsa.20451.

15 Charilaos Efthymiou. A simple algorithm for random colouring $G(n, d/n)$ using $(2 + \epsilon)d$ colours. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2012, Kyoto, Japan, January 17-19, 2012*, pages 272–280. SIAM, 2012. doi:10.1137/1.9781611973099.25.

16 Charilaos Efthymiou. Reconstruction/non-reconstruction thresholds for colourings of general galton-watson trees. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2015, August 24-26, 2015, Princeton, NJ, USA*, volume 40 of *LIPIcs*, pages 756–774. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2015. doi:10.4230/LIPIcs.APPROX-RANDOM.2015.756.

17 Charilaos Efthymiou. A simple algorithm for sampling colorings of $G(n, d/n)$ up to the gibbs uniqueness threshold. *SIAM J. Comput.*, 45(6):2087–2116, 2016. doi:10.1137/140977643.

18 Charilaos Efthymiou, Thomas P. Hayes, Daniel Stefankovic, and Eric Vigoda. Sampling random colorings of sparse random graphs. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2018, New Orleans, LA, USA, January 7-10, 2018*, pages 1759–1771. SIAM, 2018. doi:10.1137/1.9781611975031.115.

19 Uriel Feige. Relations between average case complexity and approximation complexity. In *Proceedings on 34th Annual ACM Symposium on Theory of Computing, May 19-21, 2002, Montréal, Québec, Canada*, pages 534–543. ACM, 2002. doi:10.1145/509907.509985.

20 Vitaly Feldman, Will Perkins, and Santosh S. Vempala. On the complexity of random satisfiability problems with planted solutions. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pages 77–86. ACM, 2015. doi:10.1145/2746539.2746577.

21 Alan Frieze and Eric Vigoda. A survey on the use of markov chains to randomly sample colorings. In *Combinatorics, Complexity, and Chance*, pages 53–71. Oxford University Press, January 2007. doi:10.1093/acprof:oso/9780198571278.003.0004.

22 Alan M. Frieze and Michael Anastos. Randomly coloring simple hypergraphs with fewer colors. *Inf. Process. Lett.*, 126:39–42, 2017. doi:10.1016/j.ipl.2017.06.005.

23 Andreas Galanis, Daniel Stefankovic, and Eric Vigoda. Inapproximability for antiferromagnetic spin systems in the tree nonuniqueness region. *J. ACM*, 62(6):50:1–50:60, 2015. doi:10.1145/2785964.

24 Andrei Giurgiu, Nicolas Macris, and Rüdiger Urbanke. Spatial coupling as a proof technique and three applications. *IEEE Transactions on Information Theory*, 62(10):5281–5295, 2016. doi:10.1109/TIT.2016.2539144.

25 Leslie Ann Goldberg, Russell A. Martin, and Mike Paterson. Strong spatial mixing with fewer colors for lattice graphs. *SIAM J. Comput.*, 35(2):486–517, 2005. doi:10.1137/S0097539704445470.

26 Oded Goldreich. *Candidate One-Way Functions Based on Expander Graphs*, pages 76–87. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. doi:10.1007/978-3-642-22670-0_10.

27 Thomas P. Hayes, Juan Carlos Vera, and Eric Vigoda. Randomly coloring planar graphs with fewer colors than the maximum degree. *Random Struct. Algorithms*, 47(4):731–759, 2015. doi:10.1002/rsa.20560.

28 Mark Jerrum. *Counting, Sampling and Integrating: Algorithm and Complexity*. Lectures in Mathematics ETH. Birkhäuser Basel, 1st edition, 2003.

29 Florent Krzakala, Andrea Montanari, Federico Ricci-Tersenghi, Guilhem Semerjian, and Lenka Zdeborová. Gibbs states and the set of solutions of random constraint satisfaction problems. *Proceedings of the National Academy of Sciences*, 104(25):10318–10323, 2007. doi:10.1073/pnas.0703685104.

30 Florent Krzakala and Lenka Zdeborová. Hiding quiet solutions in random constraint satisfaction problems. *Phys. Rev. Lett.*, 102:238701, June 2009. doi:10.1103/PhysRevLett.102.238701.

31 M Mezard, G Parisi, and M Virasoro. *Spin Glass Theory and Beyond*. WORLD SCIENTIFIC, 1986. doi:10.1142/0271.

32 M. Mézard, G. Parisi, and R. Zecchina. Analytic and algorithmic solution of random satisfiability problems. *Science*, 297(5582):812–815, 2002. doi:10.1126/science.1073287.

33 Marc Mézard and Andrea Montanari. *Information, Physics, and Computation*. Oxford Graduate Texts. Oxford University Press, Oxford, 2009.

34 Michael Molloy. The freezing threshold for k-colourings of a random graph. *J. ACM*, 65(2):7:1–7:62, 2018. doi:10.1145/3034781.

35 Elchanan Mossel and Allan Sly. Exact thresholds for Ising–Gibbs samplers on general graphs. *The Annals of Probability*, 41(1):294–328, 2013. doi:10.1214/11-AOP737.

36 Sheldon M Ross. *Stochastic Processes*. John Wiley & Sons, Nashville, TN, 2 edition, January 1995.

37 Daniel L Stein and Charles M Newman. *Spin Glasses and Complexity*. Princeton University Press, January 2013.

38 Michel Talagrand. The Parisi formula. *Ann. Math. (2)*, 163(1):221–263, 2006. doi:10.4007/annals.2006.163.221.

39 Eric Vigoda. Improved bounds for sampling colorings. In *40th Annual Symposium on Foundations of Computer Science, FOCS '99, 17-18 October, 1999, New York, NY, USA*, pages 51–59. IEEE Computer Society, 1999. doi:10.1109/SFFCS.1999.814577.