Polynomial-time sampling despite disorder chaos

Eric Ma ERYMA@STANFORD.EDU
Stanford

Tselil Schramm TSELIL@STANFORD.EDU
Stanford

Editors: Steve Hanneke and Tor Lattimore

Abstract

A distribution over instances of a sampling problem is said to exhibit *transport disorder chaos* if perturbing the instance by a small amount of random noise dramatically changes the stationary distribution (in Wasserstein distance). Seeking to provide evidence that some sampling tasks are hard on average, a recent line of work has demonstrated that disorder chaos is sufficient to rule out "stable" sampling algorithms, such as gradient methods and some diffusion processes.

We demonstrate that disorder chaos does not preclude polynomial-time sampling by canonical algorithms in canonical models. We show that with high probability over a random graph $\boldsymbol{G} \sim G(n, 1/2)$: (1) the hardcore model (at fugacity $\lambda = 1$) on $\boldsymbol{G}$ exhibits disorder chaos, and (2) Glauber dynamics run for $O(n)$ time can approximately sample from the hardcore model on $\boldsymbol{G}$ (in Wasserstein distance).

Keywords: sampling, statistical physics, average-case complexity

1. Introduction

Suppose we are given an instance of an optimization problem, such as the maximum independent set problem on a graph G. Instead of solving the optimization problem and returning a maximum independent set S^* in G, we may instead wish to *sample* from a distribution μ_G over solutions, where the probability of sampling an independent set S is a function of the objective value (here, size) of S. For example, the well-studied *hardcore model* with *fugacity* $\lambda > 0$ assigns to each set $S \subset V(G)$ a probability

$$\mu_G(S) \propto \lambda^{|S|} \cdot \mathbf{1}[S \text{ independent set in } G].$$

If $\lambda = 1$, μ_G is the uniform distribution over independent sets in G; on the other hand, as $\lambda \to \infty$, μ_G becomes concentrated on large independent sets, eventually converging to the uniform measure over optimal solutions S^*. This sampling problem and its cousins originate at the intersection of theoretical computer science and statistical physics, and sampling is by now mainstream in theoretical computer science.

The discussion above makes it clear that optimization reduces to sampling. For example, in the hardcore model, sampling from μ_G at sufficiently large fugacity λ gives a randomized algorithm for solving the maximum independent set problem on G. For this reason, if P $\neq$ NP, we expect that sampling in general requires super-polynomial time. Even in the average case, when the graph $\boldsymbol{G}$ is drawn from, say, the Erdős-Rényi distribution, the best polynomial-time algorithms known today can only find an independent set of size $\frac{1}{2}$ of optimal, and so even in average-case graphs, sampling from μ_G at large enough fugacity is (conjecturally) hard.

In fact, for some optimization problems it is believed that sampling is strictly harder than optimization, in that the sampling problem appears to be hard, even when the optimization problem is known to be solvable in polynomial time. One well-studied example is the max-cut problem on a graph with independent standard Gaussian weights, also known as the *Sherrington Kirkpatrick Ising model*. There, a striking result of Montanari (2021) gives a PTAS for the optimization problem, but the associated sampling problem is conjectured hard (see e.g. El Alaoui et al. (2022)). The same phenomenon occurs in several average-case sampling/optimization problems, including polynomial optimization with Gaussian coefficients ("p-spin models", see Subag (2021); El Alaoui et al. (2025a); Huang et al. (2024)), and finding a Boolean vector in an intersection of random halfspaces ("perceptron problems", see Bansal and Spencer (2020); Abbe et al. (2022); El Alaoui and Gamarnik (2024)). It is possible (though not widely believed, see discussion below) that this also occurs in the hardcore model in sparse random graphs $\boldsymbol{G} \sim G(n, d/n)$, in that polynomial-time optimization algorithms produce independent sets of size $(1 + o_d(1))\frac{\log d}{d}n$, but we know no efficient algorithm that samples from $\mu_{\boldsymbol{G}}$ at the corresponding fugacity $\lambda = \Omega(1/\log d)$.

We want to understand whether average-case sampling problems, such as sampling from the hardcore model on Erdős-Rényi graphs, are hard for polynomial-time algorithms. Complexity cannot always follow from the optimization-to-sampling reduction, as sometimes the associated optimization problem is easy. So far the leading approach is to prove lower bounds against specific algorithms. For example, we might try to show super-polynomial mixing time of the canonical *Glauber dynamics* Markov Chain, whose stationary distribution is $\mu_{\boldsymbol{G}}$. Such lower bounds are a good place to start, but they are dissatisfying in that they only rule out a very specific (albeit canonical) algorithm.

Recently, El Alaoui et al. (2022) have proposed *transport disorder chaos* as a more general criterion for hardness of sampling, and have shown that it gives unconditional lower bounds against *smooth* algorithms. A sampling problem $\mu_{\boldsymbol{G}}$ exhibits transport disorder chaos if small random perturbations $\boldsymbol{G}'$ of the instance $\boldsymbol{G}$ tend to result in a large Wasserstein (or "transport") distance between $\mu_{\boldsymbol{G}}$ and $\mu_{\boldsymbol{G}'}$ (this is qualitatively similar to, though technically weaker than, the overlap-gap property, see the survey Gamarnik (2021)). Smooth algorithms are algorithms which are robust to small perturbations: $\mathcal{A}$ is a smooth sampling algorithm if the Wasserstein distance between the output distributions $\mathcal{A}(\boldsymbol{G})$ and $\mathcal{A}(\boldsymbol{G}')$ is small. Thus, given that $\mu_{\boldsymbol{G}}$ exhibits transport disorder chaos, the triangle inequality implies that smooth algorithms cannot sample from $\mu_{\boldsymbol{G}}$ in Wasserstein distance.

El Alaoui, Montanari, and Sellke introduced this idea in the context of the Sherrington Kirkpatrick Ising model as evidence that their algorithmic method is close to tight (at least among smooth algorithms): they present a smooth sampling algorithm based on stochastic localization that works up to *inverse temperature* (the analogue of fugacity) $\beta = \frac{1}{2}$ (later improved to $\beta < 1$ in Celentano (2024)), and then show that disorder chaos onsets at larger inverse temperature $\beta > 1$. This style of lower bound has since been utilized for other problems, including p-spin models (El Alaoui et al. (2025a,b)), perceptron problems (El Alaoui and Gamarnik (2024)), and Ising models on sparse graphs (Huang et al. (2025)). We note that there are other algorithm-specific lower bounds; in the case of the Sherrington-Kirkpatrick model, there are conductance lower bounds that apply to Glauber dynamics with a worst-case initialization Sellke (2025).

Which natural sampling algorithms are smooth? In El Alaoui et al. (2022, 2025a), the authors prove that in p-spin models, an algorithm comprising of $O(1)$-steps of gradient descent or discretized stochastic localization is stable. In the context of p-spin models, these algorithms are state-of-the-art for sampling (and the disorder chaos lower bounds of El Alaoui et al. (2022, 2025a) conclusively show that their analysis of these algorithms is sharp). But for other average-case sampling problems, it is not clear whether the canonical algorithms are smooth. Further, even in p-spin models, gradient descent is not smooth if it is run for $\omega(1)$ steps[1] (though for most p-spin models, analyzing $\omega(1)$ steps seems beyond the reach of current techniques).

In a recent work, Li and Schramm (2024) begin to investigate whether disorder chaos could be a barrier to a larger class of algorithms. They demonstrate that the (not especially canonical) short (s, t)-path problem in $G(n, 2\log n/n)$ on the one hand exhibits transport disorder chaos, but on the other hand admits polynomial-time sampling algorithms. However, their sampling algorithms are based on brute-force enumeration: the corresponding measure $\mu_{\boldsymbol{G}}$ is concentrated on $\mathrm{poly}(n)$ efficiently-enumerable solutions. Most

1. The top eigenvector of a symmetric Gaussian matrix is computable by gradient descent, but is not a smooth function of the matrix.

sampling problems of interest have μ anti-concentrated on any subset of poly(n) solutions. Thus, sampling by brute-force enumeration rarely yields efficient algorithms. Adding to this that (s,t)-path has not been previously studied as a sampling problem, and Li and Schramm's counterexample might seem like a curious anomaly, without clear implications for the complexity of typical sampling problems.

The main question we wish to address in this work is whether transport disorder chaos implies hardness of sampling by canonical algorithms. We ask whether one can furnish a more canonical sampling problem which is a counterexample in that it (1) can be solved by polynomial-time algorithms, while (2) exhibiting transport disorder chaos. Indeed, we are able to show that at fugacity $\lambda = 1$, polynomial-time Glauber dynamics solves the problem of sampling from the hardcore model over $\boldsymbol{G} \sim G(n, 1/2)$ (in Wasserstein distance), even while $\mu_{\boldsymbol{G}}$ exhibits disorder chaos. Our sampling result may be of independent interest, as no efficient algorithm is known to sample from the hardcore model on $G(n, 1/2)$ at this fugacity.

1.1. Results

To state our results, we require a definition:

Definition 1 (Normalized Wasserstein/transport distance) *If μ, ν are distributions over a space equipped with ℓ_2 metric $\|\cdot\|$, the normalized 2-Wasserstein or transport distance between μ and ν is the quantity*

$$\overline{W}_2(\mu, \nu) \stackrel{\text{def}}{=} \inf_{\pi \in C(\mu,\nu)} \sqrt{\mathop{\mathbf{E}}_{(\boldsymbol{X},\boldsymbol{Y})\sim\pi} \left\| \frac{\boldsymbol{X}}{\sqrt{\mathbf{E}_{\boldsymbol{X}\sim\mu} \|\boldsymbol{X}\|^2}} - \frac{\boldsymbol{Y}}{\sqrt{\mathbf{E}_{\boldsymbol{Y}\sim\nu} \|\boldsymbol{Y}\|^2}} \right\|^2},$$

where $C(\mu, \nu)$ is the space of all couplings of μ and ν.

Remark 2 *Prior works on disorder chaos normalize by n instead of the second moment of each random variable, as we are doing. This is because they are working with distributions on $\{\pm 1\}^n$ (such as Ising models or the Sherrington-Kirkpatrick model), so the second moments are always n. Moreover, we normalize $\boldsymbol{X}, \boldsymbol{Y}$ separately because we want to measure their correlation. In our case, the moments of $\boldsymbol{X}, \boldsymbol{Y}$ concentrate well enough that our results would hold if we jointly normalized them.*

We will represent μ_G as a distribution over the ℓ_2-space $\{0,1\}^n$, associating each independent set S with its indicator vector. Our first result is that the hardcore model on $G(n, 1/2)$ exhibits disorder chaos, meaning that small perturbations of $\boldsymbol{G}$ result in large perturbations of $\mu_{\boldsymbol{G}}$, as measured in normalized Wasserstein distance:

Theorem 3 (The hardcore model exhibits disorder chaos) *Consider correlated graphs $\boldsymbol{G}, \boldsymbol{G}' \sim G(n, 1/2)$, where $\boldsymbol{G}'$ is given by re-sampling each edge in $\boldsymbol{G}$ independently with probability $s \in (0, 1)$. Let $\mu_{\boldsymbol{G}}, \mu_{\boldsymbol{G}'}$ denote the hardcore model at fugacity $\lambda = 1$ on $\boldsymbol{G}, \boldsymbol{G}'$ respectively. Then on average, even for arbitrarily small s the distributions $\mu_{\boldsymbol{G}}$ and $\mu_{\boldsymbol{G}'}$ are as separated as possible in normalized Wasserstein distance,*

$$\lim_{s\to 0} \liminf_{n\to\infty} \mathop{\mathbf{E}}_{\boldsymbol{G},\boldsymbol{G}'} \overline{W}_2(\mu_{\boldsymbol{G}}, \mu_{\boldsymbol{G}'})^2 = 2.$$

The intuition for Theorem 3 is straightforward. Since we are representing subsets of $[n]$ with their indicators in $\{0,1\}^n$, the squared distance between independent sets $\boldsymbol{S} \in \boldsymbol{G}$ and $\boldsymbol{S}' \in \boldsymbol{G}'$ is $\|\boldsymbol{S} - \boldsymbol{S}'\|^2 = |\boldsymbol{S}| + |\boldsymbol{S}'| - 2|\boldsymbol{S} \cap \boldsymbol{S}'|$. The quantity $|\boldsymbol{S} \cap \boldsymbol{S}'|$ is at most the size of the largest subset of vertices in $\boldsymbol{S}$ which form an independent set in $\boldsymbol{G}'$. Say $|\boldsymbol{S}| = m$; since each edge in $\boldsymbol{G}'$ is resampled with probability s, the graph induced on $\boldsymbol{S}$ in $\boldsymbol{G}'$ is distributed according to $G(m, s/2)$. The key fact is that in $G(m, s/2)$, the size of the maximum independent set scales as $2 \log m / \log(1 - s/2)^{-1}$ with high probability. Hence, so long as $s = \Omega_m(1)$, $|\boldsymbol{S} \cap \boldsymbol{S}'| = O(\log m) \ll m = |\boldsymbol{S}|$, and $\|\boldsymbol{S} - \boldsymbol{S}'\|^2 = |\boldsymbol{S}| + |\boldsymbol{S}'| - o(|\boldsymbol{S}|)$. At fugacity $\lambda = 1$, the sizes of $\boldsymbol{S} \sim \mu_{\boldsymbol{G}}$ and

$S' \sim \mu_{G'}$ concentrate well and grow as $n \to \infty$, which suffices to argue that $\overline{W}_2^2(\mu_G, \mu_{G'}) = 2(1 - o_n(1))$ with high probability.

We remark that the same argument can be used to prove that $\overline{W}_2(\mu_G, \mu_{G'}) = \Omega_s(1)$ when $\boldsymbol{G} \sim G(n, p)$ for any $\frac{1}{2} > p = \Omega(1/n)$ and $\lambda = \Theta_n(1)$ (though one must make adjustments to the asymptotics of $|\boldsymbol{S} \cap \boldsymbol{S}'|$ when p is small). For the sake of simplicity, we have fixed $p = \frac{1}{2}$ and fugacity $\lambda = 1$.

We also show that despite the disorder chaos, the canonical Markov Chain sampling algorithm for the hardcore model induces a distribution which is close to μ_G in $\overline{W}_2$ distance.

Theorem 4 (Polynomial-time sampling in Wasserstein distance) *Suppose $\boldsymbol{G} \sim G(n, 1/2)$ and let $\mu_{\boldsymbol{G}}$ be the hardcore model on $\boldsymbol{G}$ at fugacity $\lambda = 1$. For an integer $k \geqslant 0$, let $\mu_k^{\text{Glauber}}(\boldsymbol{G})$ be the distribution of the Glauber dynamics Markov Chain on $\boldsymbol{G}$ after initializing from the empty set, and running until the stopping time $\boldsymbol{\tau} = \min\{100(\log n)2^k), \boldsymbol{\tau}_k\}$, where $\boldsymbol{\tau}_k$ is the first time the Markov Chain hits an independent set of size k. Then for $k = \log n - 30 \log\log n$, with probability $1 - o_n(1)$,*

$$\overline{W}_2(\mu^{\boldsymbol{G}}, \mu_k^{\text{Glauber}}(\boldsymbol{G})) = o_n(1).$$

Remark 5 *For simplicity our theorem is stated and proved only at fugacity $\lambda = 1$, but similar arguments should work for any $\lambda \leqslant (\log n)^{O(1)}$. Observe that $\lambda = 1$ corresponds to independent sets of typical size from a typical graph drawn from $G(n, 1/2)$, e.g., independent sets of size approximately $\log n - \log\log n$. When $\lambda = \omega(\text{polylog}(n))$, we do not know of any algorithm that can find any independent set of typical size for such a λ, and it is conjectured that it is computationally hard to do so. If this conjecture holds, we would have no hope of sampling at such a fugacity.*

The Glauber Dynamics is a Markov Chain, whose state at time t is an independent set $\boldsymbol{S}_t$ of $\boldsymbol{G}$. At step $t + 1$, the dynamics samples a vertex $v \sim [n]$ uniformly at random. If $v \in \boldsymbol{S}_t$, the dynamics decides whether to drop v from $\boldsymbol{S}_t$ based on the outcome of a biased coinflip; otherwise if $v \notin \boldsymbol{S}_t$ but $\{v\} \cup \boldsymbol{S}_t$ forms an independent set, v is added to $\boldsymbol{S}_t$ based on the outcome of a biased coinflip.

To prove Theorem 4, we first give a simple argument that couples the trajectory of Glauber Dynamics initialized at $\boldsymbol{S}_0 = \emptyset$ up to time $\boldsymbol{\tau}_k$ with the (randomized) greedy algorithm for finding an independent set in $\boldsymbol{G}$. Concentration phenomena for nested independent sets in $\boldsymbol{G}$ allow us to show that if the greedy algorithm is run for $k = \log n - \Omega(\log\log n)$ steps, the output is close in total variation distance to the uniform distribution over k-independent-sets in $\boldsymbol{G}$. We construct a good $\overline{W}_2$ coupling by (1) sampling a k-independent set $\boldsymbol{S}$ using Glauber dynamics, (2) sampling an integer k^+ according to the marginal over $|\boldsymbol{S}'|$ for $\boldsymbol{S}' \sim \mu_{\boldsymbol{G}}$, and (3) if possible, outputting a uniformly random k^+-sized independent set $\boldsymbol{S}'$ which contains $\boldsymbol{S}$ (otherwise output an arbitrary k^+-independent set in $\boldsymbol{G}$). We argue that k^+ is concentrated in a window of size $2\log\log n$ around $\log n$, so that with high probability $k^+ > k$ and step (3) is valid. Then, we establish that for each k^+, the distribution induced by first sampling a k-independent set, and then sampling a random k^+ independent set which contains it, is close to uniform over k^+ independent set in $\boldsymbol{G}$. Our proof is not too complicated: the conclusion follows in a straightforward way from sufficiently strong concentration of the number of independent sets of size k^+ containing S in $\boldsymbol{G}$; we establish concentration by bounding $\Omega(\log\log n)$ moments via combinatorial arguments. Since $\mu_G(S)$ treats all independent sets of equal size symmetrically, and since $|S \cap S'|/|S'| = 1 - o_n(1)$, the coupling witnesses a $\overline{W}_2$ distance of $o_n(1)$.

1.2. Discussion and open problems

Together, Theorems 3 and 4 demonstrate that even for very canonical sampling problems, disorder chaos is not incompatible with the success of very canonical polynomial-time (Wasserstein) sampling algorithms. The upshot is that even when run on average-case sampling problems, many canonical sampling algorithms are simply not smooth in the sense of El Alaoui, Montanari, and Sellke. Our results demonstrate that the set of non-smooth algorithms includes Glauber dynamics in the hardcore model at some fugacities; presumably this is true of Glauber dynamics in other contexts as well.[2]

A more troubling point is that the class of non-smooth algorithms includes iterative algorithms, such as gradient methods or stochastic localization, when run for $\omega(1)$ steps. The state-of-the-art algorithms for sampling from p-spin models are such iterative methods run for $O(1)$ steps, and disorder chaos has been used as some evidence of their optimality. Though it is not really our expectation that $\text{poly}(n)$ steps of a stochastic localization procedure can sample from p-spin models in the hard regime, our results still sow some doubt on the finality of the disorder-chaos based lower bounds.

A separate question concerns sampling from the hardcore model. The best-studied average-case sampling question for the hardcore model concerns *sparse* random graphs $\boldsymbol{G} \sim G(n, d/n)$ when $d = O_n(1)$, and specifically the question of sampling in *total variation*: the goal is to have a polynomial-time algorithm $\mathcal{A}$ which with high probability satisfies $\mathrm{d_{TV}}(\mu_{\boldsymbol{G}}, \mathcal{A}(\boldsymbol{G})) = o(1)$. By analogy with the d-regular tree, it has been conjectured that Glauber Dynamics can sample in polynomial time in this setting for fugacity up to the reconstruction threshold at $\lambda = \Omega_d(1)$ (see Restrepo et al. (2014); Bhatnagar et al. (2016)). The largest fugacity at which Glauber Dynamics is currently known to sample from $\mu_{\boldsymbol{G}}$ in total variation is $\lambda = O(d^{-1})$ Bezáková et al. (2024); Efthymiou and Feng (2023); in the qualitatively similar random d-regular graph model, this is improved to $\lambda = O(d^{-1/2})$ (Chen et al. (2025)) (and the recent result of Liu et al. (2024) for the Ising models on $G(n, d/n)$ might generalize to allow $\lambda = O(d^{-1/2})$ in the hardcore model). It is difficult to make an apples-to-apples comparison between fugacity in $G(n, 1/2)$ and $G(n, d/n)$, so instead, we compare the typical size of an independent set $\boldsymbol{S} \sim \mu_{\boldsymbol{G}}$: when $\boldsymbol{G} \sim G(n, 1/2)$, our Theorem 4 shows that the Glauber dynamics can sample in Wasserstein when λ is set such that the sets $\boldsymbol{S} \sim \mu_{\boldsymbol{G}}$ have size within a factor $1/2$ of optimal. On the other hand, when $\boldsymbol{G} \sim G(n, d/n)$ or from the random d-regular graph model, the largest fugacities at which Glauber is known to sample in total variation produce sets that are only within a factor up to $1/4$ of optimal.

Given our Theorem 4, in $G(n, 1/2)$ Glauber can sample in Wasserstein distance in a regime where it is not currently known to sample in total variation distance. Of course, it may be the case that Glauber can sample in total variation up to $\lambda = \Omega(1)$ in both dense and *sparse* Erdős-Rényi graphs. Our proof strategy cannot be directly used to establish such a result because of a lack of concentration; one would have to use different arguments. But alternatively, it is also possible that the conjecture derived from the d-regular tree is incorrect, and that sampling in Wasserstein distance is strictly easier than sampling in total variation. We are intrigued to know the answer, and we wonder what evidence for this possibility could look like.

Organization

In Section 2 we discuss notation and give a few definitions and standard technical lemmas. Section 3 establishes concentration properties of the hardcore model on $G(n, 1/2)$ (specifically, that the sizes of independent sets drawn from the hardcore model are very well concentrated around a specific value k^*). In Section 4, we prove that the hardcore model on Erdős-Rényi graphs exhibits disorder chaos (Theorem 3), and in Section 5 we prove that Glauber Dynamics samples from the hardcore model in transport distance (Theorem 4).

2. Preliminaries

In this section we describe our notation conventions, give some definitions, and mention some useful standard technical lemmas.

2. We mention that Glauber dynamics is known (Ben Arous and Jagannath (2018)) to mix slowly for many p-spin models at low temperatures, though as far as we know there is no lower bound for the Sherrington Kirkpatrick model or other models with "full replica symmetry breaking."

We will follow the convention that random variables are written in bold font. We will occasionally abbreviate independent set as i.s. In this work, $\log \stackrel{\text{def}}{=} \log_2$ unless otherwise specified. We use standard big-Oh notation; at times we use the possibly-ambiguous $O(1)$; where there is room for confusion, we add a subscript to signify the variable with respect to which the quantity is bounded, say $f(m) = O_m(1)$ to signify that $\lim_{m\to\infty} f(m) < \infty$ or $f(\varepsilon) = O_\varepsilon(1)$ to signify that $\lim_{\varepsilon\to 0} f(\varepsilon) < \infty$. For a measure μ on an ℓ_2 space, we will let $m_2(\mu) = \sqrt{\mathbf{E}_{X\sim\mu}\|X\|_2^2}$. For probability measures μ, ν, let $C(\mu, \nu)$ denote the set of all couplings of μ, ν.

3. Properties of the hardcore model on an Erdős-Rényi graph

Let μ_G denote the law of the hardcore model with fugacity $\lambda = 1$ on $\boldsymbol{G} \sim G(n, 1/2)$; that is, for each subset $S \subset [n]$,

$$\mu_G(S) \propto \lambda^{|S|}\mathbf{1}[S \text{ independent set in } \boldsymbol{G}].$$

Since $\boldsymbol{G} \sim G(n, 1/2)$ is random, μ_G is a random measure.

Let $\boldsymbol{Z}$ be the partition function of μ_G, and let $\boldsymbol{Z}_k$ denote the contribution to $\boldsymbol{Z}$ from k-sized independent sets; as $\lambda = 1$, $\boldsymbol{Z}_k$ is just the number of k-independent sets. In Lemma 6, we will show that for most graphs $\boldsymbol{G} \sim G(n, 1/2)$, the size of an independent set drawn from μ_G is highly concentrated around the value $k^* := \log n - \log\log n$. We intend that k^* is an integer (the closest integer to $\log n - \log\log n$), but we suppress floor/ceiling notation where it does not affect clarity to keep notation clean.

Let $\mu_{G|(-a,a)}$ be μ_G conditioned on drawing an independent set of size in the range $[k^* - a, k^* + a]$. Let $\mu_{G|k}$ be μ_G conditioned on drawing an independent set of size k.

Lemma 6 *Let $a \leqslant k^*$. Then with probability $1 - o(1)$,*

$$\mathrm{d_{TV}}(\mu_G\,,\,\mu_{G|(-a,a)}) = O(2^{-a}) + 2^{-\Omega(\log^2 n)}.$$

Proof

The probability of sampling an independent set of size k under μ is given by the ratio $\boldsymbol{Z}_k/\boldsymbol{Z}$. We will argue that all $\boldsymbol{Z}_k$ concentrate well enough around $M_k := \mathbf{E}[\boldsymbol{Z}_k]$ for all k, and then argue that the contribution of k outside the window $k^* \pm a$ is negligible.

Let $\boldsymbol{E}$ be the event that for all k in the set $L \stackrel{\text{def}}{=} [0, 2\log n - 5\log\log n]$, $|\boldsymbol{Z}_k - M_k| < \frac{1}{100}M_k$, and that for all other $k \notin L$, $|\boldsymbol{Z}_k - M_k| < n^2 M_k$. For the larger $k \notin L$, by Markov's inequality, $\mathbf{Pr}[|\boldsymbol{Z}_k - M_k| \geqslant n^2 M_k] \leqslant \frac{1}{n^2}$. In Section B we will argue via the second moment method that stronger concentration is likely for $k \in L$:

Lemma 7 *Let $\varepsilon > 0, k \leqslant 2\log n - 5\log\log n$. Then*

$$\mathbf{Pr}\left[|\boldsymbol{Z}_k - M_k| \geqslant \varepsilon M_k\right] \leqslant O\left(\frac{k^5}{\varepsilon^2 n^2}\right).$$

Applying a union bound over k, $\mathbf{Pr}[\boldsymbol{E}] \geqslant 1 - O(\frac{\log^6 n}{n^2})$. Henceforth, we condition on $\boldsymbol{E}$.

As we will argue in Section B, the sequence M_k is well-behaved, decreasing geometrically around k^*:

Lemma 8 *Let $K > 0$ be an integer. Suppose $k = k^* + K$. Then*

$$\frac{M_{k+1}}{M_k} \leqslant \left(1 + O\left(\frac{\log\log n}{\log n}\right)\right)2^{-K}.$$

If $k = k^* - K$, then

$$\frac{M_{k-1}}{M_k} \leqslant \left(1 + O\left(\frac{\log\log n}{n}\right)\right)2^{-K-1}.$$

So for any constant factor $A < 2$, for n large enough, the sequence M_k decreases geometrically by A as k ascends/descends from k^*. Conditioned on $\boldsymbol{E}$, when $k \leqslant 2\log n - 5\log\log n$, $0.99 \leqslant Z_k/M_k \leqslant 1.01$, so up to k not too large the Z_k sequence is also geometrically decreasing around k^* by a (conservative) factor of 1.5.

It remains to argue about the independent sets of size at least $j \geqslant 2\log n - 5\log\log n$. By iteratively applying Lemma 8, we see that

$$\frac{M_j}{M_{k^*}} \leqslant \left(1 + O\left(\frac{\log\log n}{\log n}\right)\right)^{j-k^*} \prod_{i=1}^{j-k^*} 2^{-i} \leqslant 2^{O(j\log\log n/\log n) - \binom{j}{2}} \leqslant 2^{-\Omega(j^2)}.$$

Since $n^{2(j-k^*)} = 2^{O(j\log\log n)} = 2^{o(j^2)}$, under the event $\boldsymbol{E}$ we can transfer the upper bound on M_j/M_{k^*} to Z_j/Z_{k^*}:

$$\frac{Z_j}{Z_{k^*}} \leqslant 2^{-\Omega(j^2)}.$$

Putting the above together,

$$\begin{aligned}
\sum_{k=0}^{n} Z_k &= \sum_{k=k^*-a}^{k^*+a} Z_k + \sum_{k=0}^{k^*-a-1} Z_k + \sum_{k=k^*+a+1}^{2\log n - 5\log\log n} Z_k + \sum_{k=2\log n-5\log\log n+1}^{n} Z_k \\
&\leqslant \sum_{k=k^*-a}^{k^*+a} Z_k + O(2^{-a} Z_k^*) + n2^{-\Omega(\log^2 n)} Z_{k^*} \\
&\leqslant \left(1 + O\left(2^{-a} + 2^{-\Omega(\log^2 n)}\right)\right) \sum_{k=k^*-a}^{k^*+a} Z_k.
\end{aligned}$$

Let $\delta = O(2^{-a}) + 2^{-\Omega(\log^2 n)}$. Then we observe that the probability of sampling an independent set in the range $[k^* - a, k^* + a]$ is at least $1 - O(\delta)$. It follows from standard arguments (which we have recorded in Lemma 21) that $\mathrm{d}_{\mathrm{TV}}(\mu_G, \mu_{G|(-a,a)}) = O(\delta)$. ■

4. The hardcore model on Erdős-Rényi graphs exhibits disorder chaos

In this section, we will show that the hardcore model on $G(n, 1/2)$ exhibits *transport disorder chaos*, ultimately proving Theorem 3.

Transport disorder chaos is a property of a distribution over a family of measures; typically, each measure μ_G in the family is parameterized by some specific object, such as an n-vertex graph G, and the distribution $\mathcal{D}$ over these objects induces the distribution over measures. Disorder chaos further requires a natural notion of a random noise operator that one can apply to the underlying object G, where the noise operator fixes $\mathcal{D}$. Applying this noise operator to G naturally induces a new random G', and a new measure $\mu_{G'}$.

In our case, the parameterizing objects are graphs sampled from $G(n, 1/2)$, and the measure corresponding to a graph $\boldsymbol{G}$ is just $\mu_{\boldsymbol{G}}$, the hardcore model. The noise operator is the natural noise operator over the boolean cube $\{0,1\}^{\binom{[n]}{2}}$, which resamples each edge with some specified probability. Formally, let $T_{1-s}(G)$ be the distribution on graphs $\boldsymbol{G}'$ given by resampling each edge in the graph G independently with probability s. Equivalently, every edge of G is flipped with probability $s/2$. We say that the family μ_G has disorder chaos if

$$\inf_{s\in(0,1)} \liminf_{n\to\infty} \mathop{\mathbf{E}}_{\substack{G\sim G(n,1/2)\\ G'\sim T_{1-s}(G)}} \left[\overline{W}_2(\mu_G, \mu_{G'})\right] > 0.$$

Intuitively, the presence of disorder chaos means that if we apply any arbitrarily small random perturbation to G, the measure μ_G will typically experience a noticeable perturbation in $\overline{W}_2$ transport distance.

Theorem 9

$$\inf_{s\in(0,1)} \liminf_{n\to\infty} \mathop{\mathbf{E}}_{\substack{G\sim G(n,1/2)\\ G'\sim T_{1-s}(G)}} \left[\overline{W}_2(\mu_G,\mu_{G'})^2\right] = 2.$$

Proof (Sketch) Recall the normalized transport distance and our (natural) choice to represent μ_G as a measure over $\{0,1\}^n$,

$$\overline{W}_2(\mu_G,\mu_{G'})^2 = \inf_{\pi} \mathop{\mathbf{E}}_{(S,S')\sim\pi} \left\| \frac{\mathbf{1}_S}{\sqrt{\mathbf{E}_{S\sim\mu_G}\|\mathbf{1}_S\|^2}} - \frac{\mathbf{1}_{S'}}{\sqrt{\mathbf{E}_{S'\sim\mu_{G'}}\|\mathbf{1}_{S'}\|^2}} \right\|^2 = 2 - 2\sup_{\pi} \mathop{\mathbf{E}}_{(S,S')\sim\pi} \frac{|S\cap S'|}{\sqrt{\mathbf{E}_{S\sim\mu_G}|S|\cdot \mathbf{E}_{S'\sim\mu_{G'}}|S'|}}.$$

The upper bound $\mathbf{E}[\overline{W}_2(\mu_G,\mu_{G'})^2] \leqslant 2$ follows immediately from the non-negativity of set sizes. For the lower bound, the high-level argument is that for a typical $\boldsymbol{G}$, μ_G places most of its mass on independent sets of size roughly $\log n$. Resampling each edge in an independent set of size $\log n$ with constant probability $s > 0$ typically only leaves behind an independent set of size $o(\log n)$. Therefore, any coupling between μ_G and $\mu_{G'}$ will have to map most independent sets of size $\log n$ in μ_G to nearly disjoint independent sets also of size $\log n$ in $\mu_{G'}$. This argument is a bit technical to formalize, and can be found in the appendix. ■

5. Glauber Dynamics samples in transport distance

In this section we will prove Theorem 4. We'll introduce the Glauber Dynamics and show that it can be coupled with a simple greedy algorithm. Then, we'll show that the greedy algorithm is close to uniform over independent sets of size $k^- = \log n - O(\log\log n)$. Finally, we'll show that there is a good $\overline{W}_2$ coupling between the uniform distribution on independent sets of size k^- and μ_G.

5.1. Coupling the Glauber Dynamics with the RandomGreedy algorithm

Theorem 4 states that Glauber dynamics (Algorithm 1) (run until a stopping condition is reached) samples in $\overline{W}_2$ from μ_G. However, we will find it more convenient to analyze a randomized greedy algorithm (Algorithm 2, henceforth known as RandomGreedy). In this section, we state both algorithms and show that they can be coupled with high probability.

For a measure μ, recall that the generic Glauber dynamics is a Markov chain where at any state $\sigma \in \{0,1\}^n$, we pick a coordinate $\boldsymbol{i} \sim [n]$ uniformly at random, and then resample from the conditional measure of μ where we enforce that for our new sample X, $X_j = \sigma_j$ for all $j \neq i$. It is easy to check that the stationary distribution of this Markov chain is μ (e.g. see Levin and Peres (2017), chapter 3). For the hardcore model on any graph G with $\lambda = 1$, it is easy to check that Algorithm 1 run on graph G exactly corresponds to the generic Glauber dynamics with measure μ_G. Therefore, the stationary measure of Algorithm 1 run on G is precisely μ_G (although we do not ever use this fact in our analysis).

RandomGreedy builds an independent set, starting from $\emptyset$ and iteratively adding a uniformly random vertex as long as one is available to add. The only difference between this algorithm and Glauber dynamics is that Glauber dynamics samples a vertex uniformly from $[n]$ at each step and decides whether to add or remove it; removal happens so rarely that it is easy to construct a coupling of RandomGreedy and Glauber dynamics. This coupling will allow us to translate results from RandomGreedy to Glauber dynamics.

Algorithm 1 Glauber dynamics

Require: A graph G, a size s, and a time T
Ensure : FAIL or an independent set of size s
Let $S_0 = \emptyset$
for $t = 1$ **to** T **do**
  $v \sim \text{Unif}([n])$
  if $S_{t-1} \cup \{v\}$ *is not an independent set* **then**
    $S_t = S_{t-1}$
  end
  else
    $S_t \sim \text{Unif}\{\, S_{t-1} \cup \{v\}, S_{t-1} \setminus \{v\}\,\}$
  end
  if $|S_t| = s$ **then**
    return S_t
  end
end
return FAIL

Algorithm 2 RandomGreedy

Require: A graph G and size s
Ensure : FAIL or an independent set of size s
$S_0 = \emptyset$
for $i = 1$ **to** s **do**
  $X_i = \{v \in [n] \setminus S_{i-1} \mid \{v\} \cup S_{i-1}$ is an independent set in $G\}$
  if $X_i = \emptyset$ **then**
    return FAIL
  end
  $v_i \sim \text{Unif}(X_i)$
  $S_i = S_{i-1} \cup \{v_i\}$
end
return S_s

For a fixed graph G, size $s > 0$ and time $T > 0$, let $\text{Glauber}(G, s, T), \text{RandomGreedy}(G, s)$, denote the random variable corresponding to the output of Algorithm 1, Algorithm 2 respectively.

In both Glauber Dynamics and RandomGreedy, it will be important to us to understand how many vertices are not neighbors with any vertex of S_t (and thus can be added to S_t). This motivates the following definition:

Definition 10 *Given a graph $G = (V, E)$ and any $S \subseteq V$, the up-degree of S is defined as*

$$\boldsymbol{deg}^{\uparrow}(S) = |\{\, v \in V \backslash S \mid (u, v) \notin E(G) \text{ for all } u \in S\,\}|.$$

Note that the up-degree is also a function of the graph (though the notation does not explicitly reflect this).

Let $d_j = \mathbf{E}[\mathbf{deg}^{\uparrow}([j])]$. By linearity of expectation, $d_j = (n - j)2^{-j}$. A combination of Chernoff and union bounds provide simultaneous concentration of $\mathbf{deg}^{\uparrow}(S)$ around $d_{|S|}$ for all S not too large.

Lemma 11 *Let $\boldsymbol{G} \sim G(n, 1/2), C > 6$. Then with probability $1 - o(1)$, all sets $S \subseteq [n]$ of size $0, \dots, \log n - C \log\log n$ satisfy $|\boldsymbol{deg}^{\uparrow}(S) - d_{|S|}| \leqslant d_{|S|}^{2/3}$.*

We formally show a coupling between Glauber and RandomGreedy. The idea is simply that for most graphs, with high probability over the randomness of the algorithm, Glauber will not drop any nodes until it reaches an independent set of size $\log n - O(\log\log n)$. This is exactly what RandomGreedy does.

Theorem 12 *Fix $0 < s < \log n - C\log\log n, C > 6, T \geqslant 100 \cdot 2^s \cdot \log n$. With probability $1 - o(1)$ over the graph $\mathbf{G} \sim G(n, 1/2)$, there exists a coupling $\pi = \pi(\mathbf{G})$ of Glauber$(\mathbf{G}, s, T)$, RandomGreedy$(\mathbf{G}, s)$ such that*

$$\Pr_{\pi}[\mathit{Glauber}(\mathbf{G}, s) = \mathit{RandomGreedy}(\mathbf{G}, s)] = 1 - o(1).$$

5.2. RandomGreedy samples largeish independent sets in total variation

We will show that, with high probability over G, RandomGreedy can approximately sample an independent set in total variation distance from μ_G restricted to independent sets of size $\log n - C\log\log n$ for large enough C.

We explain some high-level reason that RandomGreedy samples up to size $\log n - C\log\log n$. Fix any independent set S of size s; without loss of generality suppose $S = \{1, \dots, s\}$. Using the notation of Algorithm 2, RandomGreedy outputs S if and only if the sequence $v_1, \dots, v_s$ is some permutation of the elements of S. The probability of seeing any given permutation π of the elements of S is proportional to $\prod_{i=1}^{s-1} |X_i|^{-1}$, where X_i is the set of all vertices that is not adjacent to any vertex in $\{\pi(1), \dots, \pi(i)\}$. Note that $|X_i| = \mathbf{deg}^{\uparrow}(S)$, and from Lemma 11 we have strong concentration for $\mathbf{deg}^{\uparrow}(S)$ for any set S not too large. As long as k is not too large, we have that $|X_1|, \dots, |X_{k-1}|$ all concentrate simultaneously, which is enough to prove our theorem.

Theorem 13 *Let $\mathbf{G} \sim G(n, 1/2), C > 6, k = \log n - C\log\log n$, and let $\mu_{alg} := \mu_{alg}(\mathbf{G})$ denote the probability measure associated with running the algorithm Algorithm 2 on $\mathbf{G}$ for time k. Then with probability $1 - o(1)$ over the randomness of $\mathbf{G}$, $\mathrm{d_{TV}}(\mu_{alg}, \mu_{G|k}) = o(1)$.*

Proof (Sketch) Let $C > 6$. Lemma 11 tells us that with probability $1 - o(1)$, for any $S \in [n]^{\leqslant \log n - C\log\log n}$, $|\mathbf{deg}^{\uparrow}(S) - d_{|S|}| \leqslant d_{|S|}^{2/3}$. In the remainder of the proof, we condition on this event. By the concentration of the up-degree, no matter what path RandomGreedy takes, we will output some independent set of size k. We will analyze the probability of outputting any particular k-independent set.

Let $\mathbf{S}$ be an arbitrary k-independent set in $\mathbf{G}$. To output $\mathbf{S}$, the $v_1, \dots, v_k$ in the description of the algorithm in Algorithm 2 must be a permutation of the vertices in $\mathbf{S}$. Fix a permutation and call the vertices $v_1, \dots, v_k$. We will now understand the probability of the algorithm inducing such a permutation. Since we have conditioned on concentration of the up-degrees,

$$p \stackrel{\text{def}}{=} \Pr[\text{RandomGreedy produces } v_1, \dots, v_k] \in \prod_{j=0}^{k-1} \frac{1}{d_j(1 \pm d_j^{-1/3})}.$$

We will show that p is close to $\prod_{j=0}^{k-1} d_j^{-1}$, so that the probability of reaching any k-set through any given permutation is similar. The event that we conditioned on guarantees that the multiplicative error is small enough for this to be true. ■

5.3. Sampling in transport distance

We have seen that Glauber Dynamics, run for an appropriate amount of time, will sample an approximately uniform independent set of size up to $k^- = \log n - C\log\log n$ for $C > 6$. This is not quite sufficient for sampling from μ_G in TV distance, as μ_G is concentrated on independent sets of size $k^* \stackrel{\text{def}}{=} \log n - 1 \cdot \log\log n$

(Lemma 6). Here we will show that sampling from $\mu_{G|k^-}$ in TV distance produces an approximate sample from μ_G in $\overline{W}_2$.

To show this, we will produce an ℓ_2-coupling between the uniform distribution over independent sets of size k^- and μ_G. At a high level, we will show that each k^- independent set is contained in approximately the same number of k-independent sets, for every $k \in k^* \pm \log\log n$.

We will first generalize our definition of up-degree to allow extension of an independent set by multiple vertices:

Definition 14 *Given a graph $G = (V, E)$ and $S \subseteq V$, the ℓ-up-degree of S is defined as*

$$\deg_\ell^{\uparrow}(S) = \left|\left\{ T \in \binom{V\setminus S}{\ell}; (u,v) \notin E \text{ for all } u \in S, v \in T, T \text{ is an i.s.} \right\}\right|.$$

For any ℓ-set T that satisfies the above condition, we will say that T completes S.

We will show that for any fixed constant B, when $\ell = B\log\log n$, the ℓ-up-degree of any set S of size k^- concentrates well when $\boldsymbol{G} \sim G(n, 1/2)$ for n sufficiently large.

Lemma 15 *Let $2 \leqslant B \leqslant C, C > 20$, $\ell = B\log\log n$, $k^- = \log n - C\log\log n$. Then, for any $0 \leqslant \varepsilon \leqslant 1$ and any set $S \in \binom{[n]}{k^-}$,*

$$\mathbf{Pr}[|\boldsymbol{deg}_\ell^{\uparrow}(S) - \mathbf{E}\,\boldsymbol{deg}_\ell^{\uparrow}(S)| \geqslant \varepsilon\,\mathbf{E}\,\boldsymbol{deg}_\ell^{\uparrow}(S)] \leqslant \left(\frac{1}{\varepsilon\log n}\right)^{\Omega(\log^5 n/\log\log n)}$$

where $\Omega(\cdot)$ hides a positive constant depending on B, C.

Proof (Sketch) This proof proceeds by bounding a poly-logarithmic moment of $\mathbf{deg}_\ell^{\uparrow}(S) - \mathbf{E}\,\mathbf{deg}_\ell^{\uparrow}(S)$. This involves a somewhat detailed understanding of the intersections of poly-logarithmically many independent sets of size ℓ. ■

We are now ready to prove the following theorem, which completes the proof of Theorem 4 (since the output distribution of RandomGreedy, μ_{alg}, is close in total variation to μ_G by Lemma 20, and since $\overline{W}_2$ is bounded by $O(1)$).

Theorem 16 *Recall the definition of μ_{alg} in Theorem 13. With probability $1 - o(1)$,*

$$\overline{W}_2(\mu_{alg}, \mu_G) = o(1)$$

Proof (Sketch) By Theorem 13, we know that with high probability over the graph, RandomGreedy approximately samples a uniform independent set of size, say, $k^- = \log n - 30\log\log n$. From Lemma 15, we know that with high probability, each independent set of size k^- sits inside roughly the same number of $(\log n - \log\log n) \pm \log\log n$-size independent set. Recall from Lemma 6 that with high probability, the hardcore model on $\boldsymbol{G}$ is well approximated by its restriction to sets of size $(\log n - \log\log n) \pm \log\log n$, so it suffices to find a good coupling to $\mu_{G|k}$ for each $k \in (\log n - \log\log n) \pm \log\log n$. Roughly, by the fact that RandomGreedy is uniform over k^- independent set and the fact that the ℓ-up degree concentrates, the distribution where we sample a k^- independent set according to RandomGreedy and then pick a uniformly random independent set of size k that contains it is approximately a coupling of μ_{alg} and $\mu_{G|k}$. This argument is made rigorous in the appendix. ■

Acknowledgments

We wish to thank Nima Anari and Sidhanth Mohanty for useful pointers to the sampling literature.

References

Emmanuel Abbe, Shuangping Li, and Allan Sly. Binary perceptron: efficient algorithms can find solutions in a rare well-connected cluster. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 860–873, 2022.

Noga Alon and Joel H Spencer. *The probabilistic method*. John Wiley & Sons, 2016.

Nikhil Bansal and Joel H Spencer. On-line balancing of random inputs. *Random Structures & Algorithms*, 57(4):879–891, 2020.

Gérard Ben Arous and Aukosh Jagannath. Spectral gap estimates in mean field spin glasses. *Communications in Mathematical Physics*, 361:1–52, 2018.

Ivona Bezáková, Andreas Galanis, Leslie Ann Goldberg, and Daniel Štefankovič. Fast sampling via spectral independence beyond bounded-degree graphs. *ACM Transactions on Algorithms*, 20(1):1–26, 2024.

Nayantara Bhatnagar, Allan Sly, and Prasad Tetali. Decay of correlations for the hardcore model on the d-regular random graph. *Electronic Journal of Probability*, 21, 2016.

Michael Celentano. Sudakov–Fernique post-AMP, and a new proof of the local convexity of the TAP free energy. *The Annals of Probability*, 52(3):923–954, 2024.

Xiaoyu Chen, Zejia Chen, Zongchen Chen, Yitong Yin, and Xinyuan Zhang. Rapid mixing on random regular graphs beyond uniqueness. *arXiv preprint arXiv:2504.03406*, 2025.

Charilaos Efthymiou and Weiming Feng. On the mixing time of Glauber dynamics for the hard-core and related models on $g(n, d/n)$. In *50th International Colloquium on Automata, Languages, and Programming (ICALP 2023)*, pages 54–1. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2023.

Ahmed El Alaoui and David Gamarnik. Hardness of sampling solutions from the symmetric binary perceptron. *arXiv preprint arXiv:2407.16627*, 2024.

Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Sampling from the Sherrington-Kirkpatrick Gibbs measure via algorithmic stochastic localization. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 323–334. IEEE, 2022.

Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Sampling from mean-field gibbs measures via diffusion processes. *Probability and Mathematical Physics*, 6(3):961–1022, 2025a.

Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Shattering in pure spherical spin glasses. *Communications in Mathematical Physics*, 406(5):1–36, 2025b.

David Gamarnik. The overlap gap property: A topological barrier to optimizing over random structures. *Proceedings of the National Academy of Sciences*, 118(41):e2108492118, 2021.

Brice Huang, Andrea Montanari, and Huy Tuan Pham. Sampling from spherical spin glasses in total variation via algorithmic stochastic localization. *arXiv preprint arXiv:2404.15651*, 2024.

Neng Huang, Will Perkins, and Aaron Potechin. Hardness of sampling for the anti-ferromagnetic Ising model on random graphs. In *16th Innovations in Theoretical Computer Science Conference (ITCS 2025)*, pages 61–1. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2025.

David A Levin and Yuval Peres. *Markov chains and mixing times*, volume 107. American Mathematical Soc., 2017.

Shuangping Li and Tselil Schramm. Some easy optimization problems have the overlap-gap property. *arXiv preprint arXiv:2411.01836*, 2024.

Kuikui Liu, Sidhanth Mohanty, Amit Rajaraman, and David X Wu. Fast mixing in sparse random ising models. In *2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 120–128. IEEE, 2024.

Andrea Montanari. Optimization of the Sherrington–Kirkpatrick Hamiltonian. *SIAM Journal on Computing*, pages FOCS19–1, 2021.

Irena Penev. Combinatorics and graph theory 1 & 2, 2022. Lecture Notes.

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

Mark Sellke. Exponentially slow mixing of the low temperature sk model, 2025. URL https://arxiv.org/abs/2511.22621.

Eliran Subag. Following the ground states of full-RSB spherical spin glasses. *Communications on Pure and Applied Mathematics*, 74(5):1021–1044, 2021.

Appendix A. Misc. standard results

We will find the following standard results useful.

Lemma 17 *Let $\mu = \sum_{i=1}^{k} p_i \mu_i, \nu = \sum_{i=1}^{k} p_i \nu_i$ be mixture distributions. Consider any set of couplings $\pi_i, i \in [k]$ where $\pi_i \in C(\mu_i, \nu_i)$ is a coupling of μ_i and ν_i. Then*

$$\overline{W}_2(\mu,\nu)^2 \leqslant \sum_{i=1}^{k} p_i \mathop{\mathbf{E}}_{(X,Y)\sim\pi_i} \left\| \frac{X}{\sqrt{\mathbf{E}_{X\sim\mu}\|X\|^2}} - \frac{Y}{\sqrt{\mathbf{E}_{Y\sim\nu}\|Y\|^2}} \right\|^2$$

Proof Observe that the following is a coupling of μ, ν: first pick $\boldsymbol{i} \in [k]$ according to the distribution $p_1, \dots, p_k$, and then sample $(X, Y) \sim \pi_{\boldsymbol{i}}$. The conclusion follows. ■

We also mention that our definition of $\overline{W}_2$ satisfies the triangle inequality. This follows from the fact that the standard W_2 distance satisfies the triangle inequality, and we are simply computing the standard W_2 distance on scaled measures.

Definition 18 *Let P be a transition matrix on on finite state space Ω. Let μ be some measure on Ω. We define the time-reversal P' of P as a transition matrix where*

$$P'(i,j) = \frac{P(j,i)\mu(j)}{\mu P(i)}.$$

Lemma 19 *$\mu PP' = \mu$, and $P(j,i) = 0 \implies P'(i,j) = 0$.*

Proof It is clear from the form of P' that $P(j,i) = 0 \implies P'(i,j) = 0$. Moreover,

$$\mu PP'(x) = \sum_{y \in \Omega} \mu P(y) P'(y,x) = \sum_{y \in \Omega} \mu P(y) \frac{P(x,y)\mu(x)}{\mu P(y)} = \mu(x) \sum_{y \in \Omega} P(x,y) = \mu(x).$$

■

Recall the following useful facts about TV distance.

Lemma 20 *Let μ, ν be probability measures on a finite set Ω. Then*

$$d_{TV}(\mu\,,\,\nu) = \inf_{\pi \in C(\mu,\nu)} \Pr_{(X,Y)\sim\pi}[X \neq Y].$$

Lemma 21 *Consider some probability space with probability measure* **Pr**. *Let E be some event that occurs with probability $1-\delta$ under* **Pr**. *For all small enough δ, $d_{TV}(\mathbf{Pr}[\cdot]\,,\,\mathbf{Pr}[\cdot|E]) = O(\delta)$.*

Proof For any event A,

$$\mathbf{Pr}[A|E] = \frac{\mathbf{Pr}[A,E]}{\mathbf{Pr}[E]} \leqslant \mathbf{Pr}[A](1+O(\delta)) \leqslant \mathbf{Pr}[A] + O(\delta),$$

and also,

$$\mathbf{Pr}[A|E] \geqslant \mathbf{Pr}[A,E] \geqslant \mathbf{Pr}[A] - \delta.$$

The result follows from the variational characterization of TV distance. ■

Lemma 22 *With probability $1-o(1)$, every independent set in $\mathbf{G} \sim G(n,1/2)$ is of size $O(\log n)$.*

Proof Let $k \geqslant 4\log n$. Observe that the expectation of the number of k-independent sets is

$$\binom{n}{k} 2^{-\binom{k}{2}} \leqslant 2^{k\log n - \binom{k}{2}} \leqslant 2^{k(\log n - \frac{k-1}{2})} \leqslant 2^{4\log n(-\log n + \frac{1}{2})} = 2^{-\Omega(\log^2 n)}$$

By Markov's inequality, the probability that there exists a k-clique is at most $2^{-\Omega(\log^2 n)}$. By a union bound over $4\log n \leqslant k \leqslant n$, the conclusion follows. ■

For any set $A \subseteq \{0,1\}^n$, call it *increasing* if $x \in A, y \geqslant x$ coordinate-wise implies $y \in A$. The *FKG* inequality gives a lower bound on the probability of intersections of increasing events under a class of measures known as "log-supermodular" measures that includes the uniform measure.

Lemma 23 (FKG inequality (Alon and Spencer (2016) specialized to uniform on $\{0,1\}^N$)) *Let $A_1, \dots, A_m \subseteq \{0,1\}^N$ be increasing, and* $\mathbf{Pr}[\cdot]$ *denote the uniform probability measure on $\{0,1\}^N$. Then*

$$\mathbf{Pr}[A_1 \cap \cdots \cap A_m] \geqslant \prod_{i=1}^{m} \mathbf{Pr}[A_i].$$

Lemma 24 (Penev (2022)) *Let $n \geqslant 1$. Then*

$$\frac{n^n}{e^{n-1}} \leqslant n! \leqslant \frac{n^{n+1}}{e^{n-1}}$$

Appendix B. Concentration of the number of small independent sets

In this section we will prove Lemmas 7 and 8.

Lemma *[Restatement of Lemma 7] Let $\varepsilon > 0, k \leqslant 2\log n - 5\log\log n$. Then*

$$\mathbf{Pr}\left[|\boldsymbol{Z}_k - M_k| \geqslant \varepsilon M_k\right] \leqslant O\left(\frac{k^5}{\varepsilon^2 n^2}\right).$$

Proof The proof will proceed by the second moment method. Let $\mathbf{1}_S$ be the indicator of the event that S is an independent set. Then $\boldsymbol{Z}_k = \sum_{S\in\binom{[n]}{k}} \mathbf{1}_S$. Then

$$\begin{aligned}
\mathbf{Var}\,\boldsymbol{Z}_k &= \mathbf{E}\left[\left(\sum_{S\in\binom{[n]}{k}} \mathbf{1}_S\right)^2\right] - \left(\sum_{S\in\binom{[n]}{k}} 2^{-\binom{k}{2}}\right)^2 \\
&= \sum_{S,T\in\binom{[n]}{k}} 2^{-2\binom{k}{2}+\binom{|S\cap T|}{2}} - 2^{-2\binom{k}{2}} \\
&= \sum_{r=0}^{k} \binom{n}{2k-r}\binom{2k-r}{k}\binom{k}{r}\left(2^{\binom{r}{2}} - 1\right)2^{-2\binom{k}{2}} \\
&\leqslant \sum_{r=2}^{k} \binom{n}{2k-r}\binom{2k-r}{k}\binom{k}{r}2^{\binom{r}{2}}2^{-2\binom{k}{2}},
\end{aligned}$$

where the third equality follows by counting the number of subsets S, T that intersect on r elements. Call the rth term in the summation T_r. For each $r \geqslant 2$ and n sufficiently large, we have

$$\frac{T_r}{T_2} \leqslant \frac{(2k-2)(2k-3)\cdots(2k-r)}{(n-2k+r)(n-2k+r-2)\cdots(n-2k+3)}\frac{\binom{k}{r}}{\binom{k}{2}}\frac{2^{r(r-1)/2}}{2} \leqslant \frac{2}{r!}\left(\frac{2k^2\cdot 2^{(r+1)/2}}{n-2k}\right)^{r-2} < 1,$$

where the final inequality follows because $2 \leqslant r \leqslant k \leqslant 2\log n - 5\log\log n$. Applying this to the summation above, $\mathbf{Var}\,\boldsymbol{Z}_k \leqslant k \cdot T_2$. Now,

$$\begin{aligned}
\frac{T_2}{(\mathbf{E}\,Z_k)^2} &= \frac{\binom{n}{2k-2}\binom{2k-2}{k}\binom{k}{2}2^{\binom{2}{2}}2^{-2\binom{k}{2}}}{\binom{n}{k}^2 2^{-2\binom{k}{2}}} \\
&= \frac{((n-k)!)^2(k!)^2}{(n-2k+2)!((k-2)!)^2 n!} \\
&\leqslant k^4\frac{(n-k)!^2}{n!(n-2k+2)!} = k^4\frac{(n-k)(n-k-1)\cdots(n-2k+3)}{n(n-1)\cdots(n-k+1)} = O\left(\frac{k^4}{n^2}\right)
\end{aligned}$$

where the last line follows because there are k terms in the denominator and $k-2$ terms in the numerator, where each term in the numerator is smaller than each term in the denominator. So $\mathbf{Var}\,Z_k/(\mathbf{E}\,Z_k)^2 = O(k^5/n^2)$, and the lemma follows by Chebyshev's inequality. ■

Lemma *[Restatement of Lemma 8] Let $K > 0$ be an integer. Suppose $k = k^* + K$. Then*

$$\frac{M_{k+1}}{M_k} \leqslant \left(1 + O\left(\frac{\log\log n}{\log n}\right)\right)2^{-K}.$$

If $k = k^* - K$, then

$$\frac{M_{k-1}}{M_k} \leqslant \left(1 + O\left(\frac{\log\log n}{n}\right)\right)2^{-K-1}.$$

Proof First observe that for any j,

$$\frac{M_{j+1}}{M_j} = \frac{\binom{n}{j+1}2^{-\binom{j+1}{2}}}{\binom{n}{j}2^{-\binom{j}{2}}} = \frac{n-j}{(j+1)2^j}.$$

In the first case,

$$\frac{M_{k+1}}{M_k} = \frac{(n-k-1)\log n}{(\log n - \log\log n + K)\cdot n \cdot 2^K} \leqslant 2^{-K}\frac{\log n}{\log n - \log\log n} \leqslant 2^{-K}\left(1 + O\left(\frac{\log\log n}{\log n}\right)\right).$$

In the second case,

$$\frac{M_{k-1}}{M_k} = \frac{2^{k-1}k}{n-k+1} = \frac{n}{n - \log n + \log\log n + K}\frac{\log n - \log\log n - K}{\log n}2^{-K-1} \leqslant 2^{-K-1}\left(1 + O\left(\frac{\log n}{n}\right)\right).$$

■

Appendix C. The hardcore model on Erdős-Rényi graphs exhibits disorder chaos

In the $\overline{W}_2(\mu,\nu)$ distance (Definition 1), the random variable $\boldsymbol{X} \sim \mu$ is normalized by $m_2(\mu) \stackrel{\text{def}}{=} \sqrt{\mathbf{E}_{X\sim\mu}[\|\boldsymbol{X}\|^2]}$. We require the following technical lemma, which proves that $m_2(\mu_G)$ concentrates well over the randomness of $\boldsymbol{G}$.

Lemma 25 *For any fixed constants $\alpha > 0, \beta \in (0,1)$,*

$$\Pr_{G\sim G(n,1/2)}\left[m_2(\mu_G)^2 \in (1 \pm \beta \pm o_n(1))\log n\right] \geqslant 1 - O\left(\frac{\log^6 n}{n^{2+2\alpha}}\right).$$

Proof [Proof of Lemma 25] We must establish the concentration over $\boldsymbol{G}$ of $m_2(\mu_G)^2 = \mathbf{E}_{S\sim\mu_G}\|\mathbf{1}_S\|^2 = \mathbf{E}_{S\sim\mu_G}|S|$. Recall again that $\mathbf{Pr}[|S| = k] = Z_k/Z$, and we have defined $M_k = \mathbf{E}_G[Z_k]$. By setting $\varepsilon = n^\alpha$ in Lemma 7, we see that for all $0 \leqslant k \leqslant 2\log n - 5\log\log n$,

$$\mathbf{Pr}[|Z_k - M_k| \geqslant n^\alpha M_k] \leqslant O\left(\frac{\log^5 n}{n^{2+2\alpha}}\right).$$

Furthermore, observe that $\mathbf{Pr}[|Z_k - M_k| \geqslant n^4 M_k] \leqslant 1/n^4$ by Markov's inequality for all $k > 2\log n - 5\log\log n$. Let the event that none of these events occur be E; observe that by a union bound, $\mathbf{Pr}[E] \geqslant 1 - O((\log^6 n)/n^{2+2\alpha})$. Henceforth, we condition on E.

By Lemma 8,

$$\frac{M_j}{M_{k^*}} \leqslant \left(1 + O\left(\frac{\log\log n}{\log n}\right)\right)^{j-k^*} \cdot \prod_{i=1}^{j-k^*} 2^{-i} \leqslant 2^{O(j\log\log n/\log n) - \binom{j-k^*}{2}} \leqslant 2^{-(j-k^*)^2/2 + O(j\log\log n/\log n)}.$$

so for any $k^* + \beta\log n \leqslant j \leqslant 2\log n - 5\log\log n$, we see that $M_j/M_{k^*} \leqslant 2^{-\beta^2\log^2 n/2 + O(\log\log n)}$. The same bound holds for any $j \leqslant k^* - \beta\log n$ using similar arguments. By using the same computation above, any $j > 2\log n - 5\log\log n$, $M_j/M_{k^*} \leqslant 2^{-\Omega(\log^2 n)}$. Under E, the same bounds hold in each case for Z_j/Z_{k^*}. Let $A = (k^* - \beta\log n, k^* + \beta\log n)$.

$$\Pr_{S\sim\mu_G}[|S| \notin A] = \sum_{j\notin A}\mathbf{Pr}[|S| = j] \leqslant \sum_{j\notin A} 2^{-\Omega(\beta^2\log^2 n)}\mathbf{Pr}[|S| = k^*] \leqslant 2^{-\Omega(\beta^2\log^2 n)}.$$

Our conclusion follows by an averaging argument. On the one hand,

$$m_2(\mu_G)^2 = \mathop{\mathbf{E}}_{S\sim\mu_G} |S| \leqslant (k^* + \beta \log n) + (\max_S |S|) \cdot \mathbf{Pr}[|S| > k^* + \beta n]$$
$$\leqslant k^* + \beta \log n + n \cdot 2^{-\beta^2 \log^2 n/2 + O(\log\log n)} = (1 + \beta + o_n(1)) \log n,$$

And on the other,

$$m_2(\mu_G)^2 = \mathop{\mathbf{E}}_{S\sim\mu_G} |S| \geqslant (k^* - \beta \log n)\, \mathbf{Pr}[|S| \geqslant k^* - \beta \log n]$$
$$\geqslant (k^* - \beta \log n)(1 - 2^{-\beta^2 \log^2 n + O(\log n \log\log n)}) = (1 - \beta - o_n(1)) \log n.$$

■

Theorem *[Restatement of Theorem 9]*

$$\inf_{s\in(0,1)} \liminf_{n\to\infty} \mathop{\mathbf{E}}_{\substack{G\sim G(n,1/2)\\ G'\sim T_{1-s}(G)}} \left[\overline{W}_2(\mu_G, \mu_{G'})^2\right] = 2.$$

Proof By definition of the normalized transport distance and our (natural) choice to represent μ_G as a measure over $\{0,1\}^n$,

$$\overline{W}_2(\mu_G, \mu_{G'})^2 = \inf_\pi \mathop{\mathbf{E}}_{(S,S')\sim\pi} \left\| \frac{\mathbf{1}_S}{\sqrt{\mathbf{E}_{S\sim\mu_G} \|\mathbf{1}_S\|^2}} - \frac{\mathbf{1}_{S'}}{\sqrt{\mathbf{E}_{S'\sim\mu_{G'}} \|\mathbf{1}_{S'}\|^2}} \right\|^2 = 2 - 2 \sup_\pi \mathop{\mathbf{E}}_{(S,S')\sim\pi} \frac{|S\cap S'|}{\sqrt{\mathbf{E}_{S\sim\mu_G} |S| \cdot \mathbf{E}_{S'\sim\mu_{G'}} |S'|}}.$$

The upper bound $\mathbf{E}[\overline{W}_2(\mu_G, \mu_{G'})^2] \leqslant 2$ follows immediately from the non-negativity of set sizes. For the lower bound, the high-level argument is that for a typical G, μ_G places most of its mass on independent sets of size roughly $\log n$. Resampling each edge in an independent set of size $\log n$ with constant probability $s > 0$ typically only leaves behind an independent set of size $o(\log n)$. Therefore, any coupling between μ_G and $\mu_{G'}$ will have to map most independent sets of size $\log n$ in μ_G to nearly disjoint independent sets also of size $\log n$ in $\mu_{G'}$. We now formalize this argument.

We say that a k-independent set in G "survives" resampling if it contains a sub-independent set of size at least $\log^2 k$ in $T_{1-s}G$. We define a random variable that corresponds to the fraction of k-independent sets in G that survive the resampling. Let F_k be the fraction of k-independent sets that survive resampling in G. Of course, this only makes sense if G has a k-independent set to begin with. Thus, let $F_k = 1$ if no k-independent sets exist, and otherwise let $F_k = \mathbf{E}_{G'} \mathbf{E}_{S\sim\mu_{G|k}} \mathbf{1}_{E_S}$ if G where E_S is the event that S survives (recall that $\mu_{G|k}$ is μ_G conditioned on sampling a k-set). Let A_k be the event that there is a k independent set in G.

We will now argue that

$$\mathop{\mathbf{E}}_G F_k = \mathop{\mathbf{E}}_G \left[\mathop{\mathbf{E}}_{G'} \mathop{\mathbf{E}}_{S\sim\mu_{G|k}} \mathbf{1}_{E_S} | A_k \right] \mathbf{Pr}[A_k] + \mathop{\mathbf{E}}_{bG'} [F_k | A_k^C]\, \mathbf{Pr}[A_k^C] = o(1).$$

for any $\log n - 2\log\log n \leqslant k \leqslant \log n + \log\log n$. From Lemma 7, we see that the term corresponding to A_k^C is $o(1)$. For the A_k term, observe that

$$\mathop{\mathbf{E}}_G \left[\mathop{\mathbf{E}}_{G'} \mathop{\mathbf{E}}_{S\sim\mu_{G|k}} \mathbf{1}_{E_S} | A_k \right] = \mathop{\mathbf{E}}_G \left[\mathop{\mathbf{E}}_{S\sim\mu_{G|k}} \mathop{\mathbf{E}}_{G'} \mathbf{1}_{E_S} | A_k \right]$$
$$= \mathop{\mathbf{E}}_G \left[\mathop{\mathbf{E}}_{S\sim\mu_{G|k}} \mathbf{Pr}[S \text{ survives in } G'] | A_k \right]$$

$$= \Pr_{S' \sim T_{1-s}(\overline{K_k})} [S' \text{ contains an i.s. of size at least } \log^2 k],$$

where $\overline{K_k}$ is the empty graph on k vertices. As $s > 0$ is a constant independent of n, by a union bound

$$\begin{aligned} \Pr_{S' \sim T_{1-s}\overline{K_k}} [S' \text{ contains an i.s. of size at least } \log^2 k] &\leqslant \binom{k}{\log^2 k} \cdot \left(1 - \tfrac{s}{2}\right)^{\binom{\log^2 k}{2}} \\ &\leqslant \exp\left(\log^2 k \cdot \ln k - \binom{\log^2 k}{2} \cdot \ln \frac{1}{1 - \frac{s}{2}}\right) \\ &= \exp(-\Omega(\log^4 k)) = o(1). \end{aligned}$$

Thus $\mathbf{E}_G \, F_k = o(1)$.

We now argue that $\mu_G, \mu_{G'}$ concentrate sufficiently well on $\log n$-sized sets, and that those sets typically do not survive. Let $\gamma, \delta, \beta > 0$ be small constants. From Lemma 6, we know that for any $a = O_n(1)$ with probability $1 - o(1)$,

$$\mathrm{d}_{\mathrm{TV}}(\mu_G \,,\, \mu_{G|(-a,a)}) = O(2^{-a}) + 2^{-\Omega(\log^2 n)}.$$

Thus, by setting a to be a large enough constant, we find that $\mathbf{Pr}_{S \sim \mu_G}[|S| \in [k^* - a, k^* + a]] \geqslant 1 - \gamma$. Let E be the event that this holds in both μ_G and $\mu_{G'}$. Now, let the event E' be the event that $F_k < \delta$ for all $k \in [k^* - a, k^* + a]$. By applying Markov's inequality for each k and then a union bound, $\mathbf{Pr}[E'] = 1 - o(1)$. Also, let E'' be the event that $\mathbf{E}_{S \sim \mu_G} \|S\|_2^2, \mathbf{E}_{S' \sim \mu_{G'}} \|S'\|_2^2 \in (1 \pm \beta) \log n$. By Lemma 25, $\mathbf{Pr}[E''] = 1 - o(1)$. Finally, let E''' be the event that G, G' have cliques only of size $O(\log n)$; by Lemma 22, this happens with probability $1 - o(1)$. Thus, $\mathbf{Pr}[E \cap E' \cap E'' \cap E'''] = 1 - o(1)$.

Condition on $E \cap E' \cap E'' \cap E'''$. We will now analyze any possible coupling $\pi = \pi(G, G')$ of $\mu_G, \mu_{G'}$. Let $L = [k^* - a, k^* + a]$. Observe that $1 - \gamma$ of the proportion of independent sets of G and G' have size in L. Moreover, a $(1-\gamma)(1-\delta)$ proportion of the independent sets in G have both size in L and no independent set of more than size $(1 + \beta)(\log \log n)^2$ contained inside it. Let this set of independent sets in G be $\mathcal{S}$. As just argued, $\mathcal{S}$ has probability mass at least $(1-\gamma)(1-\delta)$; π can assign at most γ of the mass of $\mathcal{S}$ to independent sets in G' with size outside of L, so $(1 - \gamma)(1 - \delta) - \gamma$ of the mass in $\mathcal{S}$ is on independent sets of size in L in $\mu_{G'}$. Thus, $1 - \eta \stackrel{\text{def}}{=} (1 - \gamma)(1 - \delta) - \gamma$ of the mass of independent sets in $\mathcal{S}$ is coupled to independent sets in G' with at most $(1 + \beta)(\log \log n)^2$ shared nodes. Therefore,

$$\overline{W}_2(\mu_G, \mu_{G'})^2 \geqslant 2 - 2\frac{(1 - \eta) \cdot (1 + \beta)(\log \log n)^2 + \eta \cdot O(\log n)}{(1 \pm \beta) \log n} \geqslant 2 - o(1) - \frac{\eta}{1 + \beta}$$

The above inequality holds for the expectation as well because $\mathbf{Pr}[E \cap E' \cap E''] = 1 - o(1)$. Thus,

$$\liminf_{n \to \infty} \mathop{\mathbf{E}}_{\substack{G \sim G(n,1/2) \\ G' \sim T_{1-s}(G)}} [\overline{W}_2^2(\mu_G, \mu_{G'})] \geqslant 2 - \frac{\eta}{1 + \beta}.$$

Since this is true for every small enough $\gamma, \delta, \beta > 0$, we can take η arbitrarily small. As the above applies to any constant $s > 0$, the conclusion follows. ■

Appendix D. Glauber dynamics samples in transport distance

D.1. Coupling

Theorem *[Restatement of Theorem 12]* *Fix $0 < s < \log n - C \log \log n$, $C > 6$, $T \geqslant 100 \cdot 2^s \cdot \log n$. With probability $1 - o(1)$ over the graph $G \sim G(n, 1/2)$, there exists a coupling $\pi = \pi(G)$ of Glauber(G, s, T), RandomGreedy(G, s) such that*

$$\Pr_{\pi}[\mathit{Glauber}(G, s) = \mathit{RandomGreedy}(G, s)] = 1 - o(1).$$

Proof First, a piece of notation: recall from the description of Algorithm 1 that the state at iteration t is the set $\mathbf{S}_t$. Let the stopping time

$$\mathbf{T}_i = \inf_{t \geqslant 0} \{ \mid \mathbf{S}_t| = i \}$$

denote the first time that Glauber dynamics reaches a set of size i.

We will now provide intuition for why such a coupling is possible. Glauber dynamics chooses a vertex i uniformly from $[n]$ at each step; if the vertex is in the independent set it is (stochastically) dropped, and if it is not in the independent set then it is (stochastically) added so long as it extends the independent set. The reason we are able to couple Glauber dynamics and RandomGreedy is because so long as the independent set has size $\leqslant s = \log n - 2 \log\log n$, there are many more vertices that could be added to the independent set than vertices within the independent set itself, so it is unlikely that Glauber dynamics drops vertices before s are added. Moreover, because it does not take too long to add a vertex at any step, with high probability, Glauber finds a set of size s in time T as above.

With the above in mind, we will construct our coupling. For $i = 0, \ldots, s$, let E_i be the event that for all $t \in [\mathbf{T}_{i-1}, \mathbf{T}_i)$, $|\mathbf{S}_t| = i - 1$. Thus, if $\cap_{i=1}^{s} E_i$ occurs, then up to reaching size s, Glauber dynamics does not ever drop any vertices. Let D_i be the event that beginning from $\mathbf{T}_{i-1}$, the time it takes for Glauber to add or drop a vertex is at most $(10 \log n)n/d_i$; therefore, if $\cap_{i=1}^{s} E_i$ and $\cap_{i=1}^{s} D_i$ occurs, it takes at most

$$\mathbf{T}_s \leqslant \sum_{i=0}^{s-1} \mathbf{T}_{i+1} - \mathbf{T}_i \leqslant 10 \log n \sum_{i=0}^{s-1} \frac{n}{n-i} 2^i \leqslant 30 \cdot 2^s \cdot \log n$$

steps of Glauber dynamics to reach an independent set of size s. Our coupling will then consist of running RandomGreedy as usual, where we let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_s$ be the sequence of vertices it adds. If $(\cap_{i=1}^{s} E_i) \cap (\cap_{i=1}^{s} D_i)$ occurs, which we claim happens with probability $1 - O(\log^{2-C} n)$, we let the vertex added at time $\mathbf{T}_i$ be $\boldsymbol{v}_i$, so that the outputs of RandomGreedy and Glauber dynamics are the same. If it does not occur, we just run Glauber dynamics and RandomGreedy as usual independently of each other. Observe that each $\boldsymbol{v}_i$ is chosen uniformly at random from the set of nodes that form an independent set with $\{ \boldsymbol{v}_1, \ldots, \boldsymbol{v}_{i-1} \}$, so this coupling results in Glauber dynamics being distributed as it should be.

We now justify that $\mathbf{Pr}[(\cap_{i=1}^{s} E_i) \cap (\cap_{i=1}^{s} D_i)] = 1 - O(\log^{2-C} n)$ with probability $1 - o(1)$ over the randomness of $\boldsymbol{G}$. First, we condition on Lemma 11, obtaining that all sets $S \in [n]^{\leqslant \log n - C \log\log n}$ satisfy $|\mathbf{deg}^{\uparrow}(S) - d_{|S|}| \leqslant d_{|S|}^{2/3}$. This occurs with probability $1 - o(1)$ over the graph $\boldsymbol{G}$. Because there exist independent sets of size s, $\mathbf{T}_0, \ldots, \mathbf{T}_s < \infty$ almost surely. We are interested in showing that $\mathbf{Pr}[E_{i+1}^C | \mathbf{S}_{\mathbf{T}_i}]$ is close to 0. This is the event that starting from a set of size i, we drop a vertex before adding one. The number of vertices that could be added is $\mathbf{deg}^{\uparrow}(S_{T_i}) \geqslant d_i(1 - d_i^{-1/3})$, where $d_i = (n - i)2^{-i}$ and the number of vertices that we can drop is i. Therefore, the probability that we drop a vertex before adding one is at most

$$i/(i + d_i(1 - d_i^{-1/3})) \leqslant 2i/d_i \leqslant 4 \log^{1-C} n$$

where the last inequality follows because i/d_i is increasing in i.

To see that $\mathbf{Pr}[D_{i+1}^C | \mathbf{S}_{\mathbf{T}_i}]$ is small, observe that we are just interested in the probability that $(10 \log n)n/d_i$ i.i.d. coins with heads probability at least $d_i(1 - d_i^{-1/3})/n$ are all tails. This is just

$$\mathbf{Pr}[D_{i+1}^C | \mathbf{S}_{\mathbf{T}_i}] \leqslant \left(1 - \frac{d_i(1 - d_i^{-1/3})}{n}\right)^{(10 \log n)n/d_i} \leqslant n^{-9}.$$

By a union bound, it is clear that $\mathbf{Pr}[(\cup_{i=1}^{s} E_i^C) \cup (\cup_{i=1}^{s} D_i^C)] \leqslant 5 \log^{2-C} n$. ■

D.2. Sampling largeish cliques

Theorem *[Restatement of Theorem 13]* *Let* $\boldsymbol{G} \sim G(n, 1/2), C > 6, k = \log n - C \log\log n$, *and let* $\mu_{alg} := \mu_{alg}(\boldsymbol{G})$ *denote the probability measure associated with running the algorithm Algorithm 2 on* $\boldsymbol{G}$ *for time* k*. Then with probability* $1 - o(1)$ *over the randomness of* $\boldsymbol{G}$, $\mathrm{d_{TV}}(\mu_{alg}, \mu_{\boldsymbol{G}|k}) = o(1)$.

Proof Let $C > 6$. Lemma 11 tells us that with probability $1 - o(1)$, for any $S \in [n]^{\leqslant \log n - C\log\log n}$, $|\mathbf{deg}^{\uparrow}(S) - d_{|S|}| \leqslant d_{|S|}^{2/3}$. In the remainder of the proof, we condition on this event. By the concentration of the up-degree, no matter what path RandomGreedy takes, we will output some independent set of size k. We will analyze the probability of outputting any particular k-independent set.

Let $\boldsymbol{S}$ be an arbitrary k-independent set in $\boldsymbol{G}$. To output $\boldsymbol{S}$, the $v_1, \dots, v_k$ in the description of the algorithm in Algorithm 2 must be a permutation of the vertices in $\boldsymbol{S}$. Fix a permutation and call the vertices $v_1, \dots, v_k$. We will now understand the probability of the algorithm inducing such a permutation. Since we have conditioned on concentration of the up-degrees,

$$p \stackrel{\text{def}}{=} \mathbf{Pr}[\text{RandomGreedy produces } v_1, \dots, v_k] \in \prod_{j=0}^{k-1} \frac{1}{d_j(1 \pm d_j^{-1/3})}.$$

We will now show that p is close to $\prod_{j=0}^{k-1} d_j^{-1}$, so that the probability of reaching any k-set through any given permutation is similar. Let $D = \prod_{j=0}^{k-1} d_j$. Observe that the possible multiplicative error of p from $1/D$ is bounded as

$$\prod_{j=0}^{k-1} \frac{1}{1 + d_j^{-1/3}} \leqslant \prod_{j=0}^{k-1} \frac{1}{1 \pm d_j^{-1/3}} \leqslant \prod_{j=0}^{k-1} \frac{1}{1 - d_j^{-1/3}}.$$

There is a negative neighborhood of 0 such that $1/(1 + x) \leqslant e^{-2x}$. Then

$$\begin{aligned}
\prod_{j=0}^{k-1} \frac{1}{1 - d_j^{-1/3}} &\leqslant \exp\left(2 \sum_{j=0}^{k-1} d_j^{-1/3}\right) = \exp\left(2 \sum_{j=0}^{k-1} \frac{2^{j/3}}{(n-j)^{1/3}}\right) \\
&\leqslant \exp\left(\frac{3}{n^{1/3}} \sum_{j=0}^{k-1} 2^{j/3}\right) \leqslant \exp\left(\frac{3}{n^{1/3}} O(2^{k/3})\right) = \exp\left(O\left(\frac{1}{\log^{C/3} n}\right)\right) \\
&= 1 + O\left(\frac{1}{\log^{C/3} n}\right).
\end{aligned}$$

By a very similar argument, we see that

$$\prod_{j=0}^{k-1} \frac{1}{1 + d_j^{-1/3}} \geqslant 1 - O\left(\frac{1}{\log^{C/3} n}\right).$$

So $p \in (1 \pm o(1))D^{-1}$. By summing over all possible permutations of S, $\boldsymbol{p}_S \stackrel{\text{def}}{=} \mathbf{Pr}[S] \in (1 \pm o(1))k!/D$. Because the algorithm terminates successfully,

$$1 = \sum_{S \text{ i.s. in } G} \boldsymbol{p}_S = Z_k \cdot (1 \pm o(1)) \frac{k!}{D},$$

Thus $1/Z_k = (1 \pm o(1))k!/D = (1 \pm o(1))\boldsymbol{p}_S$ for all independent set S, implying $\mathrm{d_{TV}}(\mu_{\mathrm{alg}}, \mu_{\boldsymbol{G}|k}) = o(1)$ as desired. ■

D.3. Transport distance coupling

Lemma *[Restatement of Lemma 15] Let $2 \leqslant B \leqslant C, C > 20$, $\ell = B \log\log n$, $k^- = \log n - C \log\log n$. Then, for any $0 \leqslant \varepsilon \leqslant 1$ and any set $S \in \binom{[n]}{k^-}$,*

$$\mathbf{Pr}[|\boldsymbol{deg}^{\uparrow}_{\ell}(S) - \mathbf{E}\,\boldsymbol{deg}^{\uparrow}_{\ell}(S)| \geqslant \varepsilon\,\mathbf{E}\,\boldsymbol{deg}^{\uparrow}_{\ell}(S)] \leqslant \left(\frac{1}{\varepsilon \log n}\right)^{\Omega(\log^5 n/\log\log n)}$$

where $\Omega(\cdot)$ hides a positive constant depending on B, C.

Proof Let $m = n - k^-$, the number of vertices remaining in G after a k^--set is removed. Fix some k^--set S; without loss of generality, let $S = \{m+1, \dots, n\}$. Let $\boldsymbol{X} = \mathbf{deg}^{\uparrow}_{\ell}(S)$, and let $\mu = \mathbf{E}\,\boldsymbol{X}$. For any even d, by Markov's inequality,

$$\mathbf{Pr}[|\boldsymbol{X} - \mu| > \varepsilon\mu] \leqslant \frac{\mathbf{E}[(\boldsymbol{X}-\mu)^d]}{\varepsilon^d \mu^d}$$

We first compute a lower bound for μ:

$$\mu = \binom{m}{\ell} 2^{-\ell k^- - \binom{\ell}{2}} = \frac{m^\ell}{\ell!} 2^{-\ell k^- - \binom{\ell}{2}} \prod_{j=0}^{\ell-1}\left(1 - \frac{j}{m}\right) \geqslant \frac{m^\ell}{\ell!} 2^{-\ell k^- - \binom{\ell}{2}} e^{-\sum_{j=0}^{\ell-1} \frac{j/m}{1-j/m}} \geqslant \frac{m^\ell}{\ell!} 2^{-\ell k^- - \binom{\ell}{2}} e^{-\frac{2}{m}\binom{\ell}{2}}$$

because $1 - x \geqslant e^{-x/(1-x)}$ for $x \in [0,1]$ and $\ell = o(m)$.

For any ℓ-set A, let $\mathbf{1}_A := \mathbf{1}[A \text{ completes } S]$. Recall that $\boldsymbol{X} = \sum_{A \in \binom{[m]}{\ell}} \mathbf{1}_A$. Let $p = \mathbf{E}\,\mathbf{1}_A$. Then

$$\mathbf{E}[(X-\mu)^d] = \mathbf{E}\left[\sum_{A_1,\dots,A_d \in \binom{[m]}{\ell}} \prod_{i=1}^{d} (\mathbf{1}_{A_i} - p)\right] = \sum_{A_1,\dots,A_d \in \binom{[m]}{\ell}} \mathbf{E}\left[\prod_{i=1}^{d} (\mathbf{1}_{A_i} - p)\right],$$

Observe that $\mathbf{E}[\mathbf{1}_A - p] = 0$. Therefore, for any choice of $A_1, \dots, A_d$, if any one of the A_is is disjoint from the rest, by independence, $\mathbf{E}[\prod(\mathbf{1}_{A_i} - p)] = 0$. By a simple counting argument (Lemma 26), if $|A_1 \cup \dots \cup A_d| > d\ell - d/2$, then such a disjoint set exists. Thus we only need to consider $A_1, \dots, A_d$ such that $|A_1 \cup \dots \cup A_d| \leqslant d\ell - d/2$.

Now, fixing any $A_1, \dots, A_d$, we claim that

$$\mathbf{E}\left[\prod_{i=1}^{d} (\mathbf{1}_{A_i} - p)\right] = 2^d\,\mathbf{Pr}[A_1, \dots, A_d \text{ complete } S].$$

This is because there are 2^d terms in the expansion of the LHS, half of which are negative. We can bound each positive product of probabilities above by $\mathbf{Pr}[\cap_{i=1}^d A_i]$ by Lemma 23 because each A_i is an increasing event.

It then follows that

$$\mathbf{E}[(\boldsymbol{X}-\mu)^d] \leqslant 2^d \sum_{a=d/2}^{d\ell-\ell} N_a \left(\frac{1}{2}\right)^{(d\ell-a)k^- + d\binom{\ell}{2} - \min\{\frac{a}{\ell}\cdot\binom{\ell}{2}, \binom{a}{2}\}},$$

where N_a is the number of possible ways to choose $A_1, \dots, A_d \in \binom{[m]}{\ell}$ such that $|A_1 \cup \dots \cup A_d| = d\ell - a$. This is because each of the $d\ell - a$ vertices in the union need to be non-neighbors of every vertex in S, and every "internal" edge within $A_1, \dots, A_d$ needs to not exist. In Lemma 27, we argue that there are at least $d\binom{\ell}{2} - \min\{\frac{a}{\ell}\cdot\binom{\ell}{2}, \binom{a}{2}\}$ of these internal edges. In Lemma 28, we prove that $N_a \leqslant \binom{m}{d\ell-a}\binom{d\ell-1}{a-1}\frac{(d\ell)!}{(\ell!)^d}$.

We now combine the upper and lower bounds with Markov's inequality:

$$\begin{aligned}
\frac{\mathbf{E}[(X-\mu)^d]}{\mu^d} &\leqslant \frac{2^d \sum_{a=d/2}^{d\ell-\ell} \binom{m}{d\ell-a}\binom{d\ell-1}{a-1}\frac{(d\ell)!}{(\ell!)^d}\left(\frac{1}{2}\right)^{(d\ell-a)k^-+d\binom{\ell}{2}-\min\{\frac{a}{\ell}\binom{\ell}{2},\binom{a}{2}\}}}{\frac{m^{d\ell}}{\ell!^d}\left(\frac{1}{2}\right)^{d\ell k^-+d\binom{\ell}{2}} e^{-\frac{2d}{m}\binom{\ell}{2}}} \\
&= 2^d e^{\frac{2d}{m}\binom{\ell}{2}} \sum_{a=d/2}^{d\ell-\ell} \frac{1}{m^{d\ell}} \binom{m}{d\ell-a}\binom{d\ell-1}{a-1}(d\ell)! 2^{ak^-+\min\{\frac{a}{\ell}\binom{\ell}{2},\binom{a}{2}\}} \\
&\leqslant 2^d e^{\frac{2d}{m}\binom{\ell}{2}} \sum_{a=d/2}^{d\ell-\ell} \frac{1}{m^{d\ell}} \frac{m^{d\ell-a}}{(d\ell-a)!}\frac{(d\ell-1)^{a-1}}{(a-1)!}\frac{(d\ell)^{d\ell+1}}{e^{d\ell-1}} 2^{ak^-+\min\{\frac{a}{\ell}\binom{\ell}{2},\binom{a}{2}\}},
\end{aligned} \tag{1}$$

where in the final line we have applied Stirling's inequality and the upper bound $x!/(x-y)! \leqslant x^y$ as well as Lemma 24. We can bound the ratio:

$$\begin{aligned}
\frac{(d\ell)^{d\ell+1}}{(d\ell-a)!} &\leqslant \frac{(d\ell)^{d\ell+1}e^{d\ell-a}}{(d\ell-a)^{d\ell-a}} = e^{d\ell-a}(d\ell)^{a+1}\left(\frac{d\ell}{d\ell-a}\right)^{d\ell-a} \\
&= e^{d\ell-a}(d\ell)^{a+1}\left(1+\frac{a}{d\ell-a}\right)^{d\ell-a} \leqslant e^{d\ell-a}(d\ell)^{a+1}e^a = e^{d\ell}(d\ell)^{a+1}.
\end{aligned}$$

Continuing from Equation (1) above and applying Stirling's inequality and dropping the $(a-1)!$ term,

$$\begin{aligned}
\frac{\mathbf{E}[(X-\mu)^d]}{\mu^d} &\leqslant 2^d e^{\frac{2d}{m}\binom{\ell}{2}} \sum_{a=d/2}^{d\ell-\ell} m^{-a}(d\ell-1)^{a-1}e^{-d\ell+1}e^{d\ell}(d\ell)^{a+1}2^{ak^-+\min\{\frac{a}{\ell}\binom{\ell}{2},\binom{a}{2}\}} \\
&\leqslant 2^d e^{\frac{2d}{m}\binom{\ell}{2}+1} \sum_{a=d/2}^{d\ell-\ell} m^{-a}(d\ell)^{2a}2^{ak^-+\min\{\frac{a}{\ell}\binom{\ell}{2},\binom{a}{2}\}} \\
&= 2^d e^{\frac{2d}{m}\binom{\ell}{2}+1} \sum_{a=d/2}^{d\ell-\ell} \left(\frac{(d\ell)^2 2^{k^-+\min\{\frac{1}{\ell}\binom{\ell}{2},\frac{a-1}{2}\}}}{m}\right)^a
\end{aligned}$$

Now using that $k^- = \log n - C\log\log n$, $\ell = B\log\log n$, and letting $d = \log^D n/\ell$,

$$\begin{aligned}
&\leqslant e2^d(1+o(1)) \sum_{a=d/2}^{d\ell-\ell} \left(\frac{n}{m}\cdot 2^{\frac{\ell-1}{2}}\log^{2D-C} n\right)^a \\
&\leqslant e2^d(1+o(1)) \sum_{a=d/2}^{d\ell-\ell} \left(\left(1+O\left(\frac{\log n}{n}\right)\right)\cdot \log^{2D-C+B/2} n\right)^a
\end{aligned}$$

So long as $2D + B/2 - C < 0$ and is bounded away from 0, the summation is geometrically decreasing, dominated by the term $a = d/2$. Thus,

$$\begin{aligned}
&\leqslant (1+o(1))e2^d\left(e\log^{2D-C+B/2} n\right)^{d/2} \\
&\leqslant (1+o(1))e\left(4e\log^{2D-C+B/2} n\right)^{d/2} \\
&\leqslant \exp\left(-F\log^D n + O(\log^D n/\log\log n)\right).
\end{aligned}$$

where $F > 0$ is some constant that depends on B, C, D. The result follows by letting $D = 5$, after which the conditions $B \leqslant C$ and $C > 20$ ensure that $2D + B/2 - C$ is strictly negative. ■

Theorem *[Restatement of Theorem 16] Recall the definition of μ_{alg} in Theorem 13. With probability $1 - o(1)$,*

$$\overline{W}_2(\mu_{alg}, \mu_G) = o(1)$$

Proof We condition on the event that every independent set in $\boldsymbol{G}$ is of size $O(\log n)$ from Lemma 22, the event that $\mathrm{d_{TV}}(\mu_G\,,\,\mu_{G|(-a,a)}) = o(1)$ for $a = \log\log n$ from Lemma 6, and the event that RandomGreedy run for a specific number of steps succeeds at sampling independent sets of size $k^- = \log n - C\log\log n$ from Theorem 13, where we choose C to be an arbitrary constant satisfying $C > 20$.

We will construct a coupling between μ_{alg} and $\mu_{G|(-a,a)}$ where we let $a = \log\log n$; this will be enough because $\mathrm{d_{TV}}(\mu_G\,,\,\mu_{G|(-a,a)}) = O(2^{-a})$. Observe that by Lemma 17, it is enough to construct a good coupling between μ_{alg} and $\mu_{G|k}$ for each k satisfying $|k - k^*| \leqslant a$. From now on, fix one such k. To couple μ_{alg} and $\mu_{G|k}$, we will sample a k^--independent set from μ_{alg}, and then choose one of the k-independent sets which extend it uniformly at random. Since $k - k^- = o(k)$, the expected distance between independent sets in this coupling will be small. It remains to verify that this induces $\mu_{G|k}$ (or something close to it) on k-independent sets.

In order to understand the relationship between the k^- and k size independent sets, consider the bipartite graph $\boldsymbol{H} = (L, R, E)$ where L consists of the independent sets of size k^-, R consists of the independent sets of size k, and $(S, T) \in E$ iff the k^- independent set S is contained in the k-independent set T. Let P be the transition operator of the simple random walk on $\boldsymbol{H}$. It is well known that one stationary measure $\mu^{\boldsymbol{H}}$ of P assigns to $v \in L \cup R$ probability proportional to its degree in $\boldsymbol{H}$. Note that the degree of a k^--independent set S in $\boldsymbol{H}$ is precisely $\mathbf{deg}_\ell^\uparrow(S)$. Let $\mu_L^{\boldsymbol{H}}, \mu_R^{\boldsymbol{H}}$ be the measure conditioned on being in L or R, respectively.

We apply Lemma 15 and a union bound to argue that the degree of every k^--independent set in L will be close to its expectation. Indeed, observe that if we let $k^- = \log n - C\log\log n$, then $\ell \stackrel{\mathrm{def}}{=} k - k^- \in [(C-2)\log\log n, C\log\log n]$, and thus ℓ satisfies the conditions of Lemma 15. We can then conclude that for any set $S \subseteq [n]$ such that $|S| = k^-$ and $\varepsilon \in (0, 1)$,

$$\mathbf{Pr}[|\mathbf{deg}_\ell^\uparrow(S) - \mathbf{E}\,\mathbf{deg}_\ell^\uparrow(S)| \geqslant \varepsilon\,\mathbf{E}\,\mathbf{deg}_\ell^\uparrow(S)] \leqslant \left(\frac{1}{\varepsilon\log n}\right)^{\Omega(\log^5 n/\log\log n)}.$$

Choosing $\varepsilon = \frac{1}{\log\log n}$ and taking a union bound over all k^--sets in $[n]$,

$$\begin{aligned}\mathbf{Pr}\left[\exists S \in \binom{[n]}{k^-}\ s.t.\, |\mathbf{deg}_\ell^\uparrow(S) - \mathbf{E}\,\mathbf{deg}_\ell^\uparrow(S)| \geqslant \frac{\mathbf{E}\,\mathbf{deg}_\ell^\uparrow(S)}{\log\log n}\right] &\leqslant \binom{n}{k^-}\left(\frac{\log\log n}{\log n}\right)^{\Omega(\log^5 n/\log\log n)} \\ &\leqslant n^{\log n}\left(\frac{\log\log n}{\log n}\right)^{\Omega(\log^5 n/\log\log n)} \\ &\leqslant \exp(-\Omega(\log^5 n)).\end{aligned}$$

Conditioning on the event that the up-degrees of all k^--sets concentrate, we have that the degree of every k^--independent set $S \in L$ is within a $(1 \pm \varepsilon)$ factor of its expectation, with $\varepsilon = 1/\log\log n$.

We now bound

$$\mathrm{d_{TV}}(\mu_L^{\boldsymbol{H}}\,,\,\mu_{\mathrm{hc},k^-}) = \sum_{S \text{ i.s. in } G, |S|=k^-} \left|\frac{1}{Z_{k^-}} - \mu_L^{\boldsymbol{H}}(S)\right|.$$

Letting D be $D = \mathbf{E}\,\mathbf{deg}_\ell^\uparrow(S)$ (by symmetry the same for all S),

$$\mu_L^{\boldsymbol{H}}(S) = \frac{\mathbf{deg}_\ell^\uparrow(S)}{\sum_{S' \in L}\mathbf{deg}_\ell^\uparrow(S')} \in \frac{(1 \pm \varepsilon)D}{Z_{k^-}(1 \pm \varepsilon)D} \subseteq (1 \pm O(\varepsilon))\frac{1}{Z_{k^-}}$$

so

$$\mathrm{d_{TV}}(\mu_L\,,\,\mu_{\mathrm{hc},k^-}) = O(\varepsilon) = o(1).$$

Recall that $\mu_L^{\boldsymbol{H}}P = \mu_R^{\boldsymbol{H}} = \mu_{G|k}$, where the last equality is because the $\boldsymbol{H}$-degree of every $T \in R$ is the same, $\binom{k}{k^-}$. Putting this together,

$$\mathrm{d_{TV}}(\mu_{\mathrm{alg}}P\,,\,\mu_{G|k}) = \mathrm{d_{TV}}(\mu_{\mathrm{alg}}P\,,\,\mu_L^{\boldsymbol{H}}P) \leqslant \mathrm{d_{TV}}(\mu_{\mathrm{alg}}\,,\,\mu_L^{\boldsymbol{H}}) \leqslant \mathrm{d_{TV}}(\mu_{\mathrm{alg}}\,,\,\mu_{\mathrm{hc},k^-}) + \mathrm{d_{TV}}(\mu_{\mathrm{hc},k^-}\,,\,\mu_L^{\boldsymbol{H}}) \leqslant o(1)$$

where the first inequality follows from the data processing inequality, and we invoke the success of RandomGreedy (Theorem 13) in the last inequality.

By Lemma 20, there exists a coupling $(\boldsymbol{X}', \boldsymbol{Y})$ of $\mu_{\mathrm{alg}}P, \mu_{G|k}$ that agrees with $1 - o(1)$ probability. It remains to find a coupling π of μ_{alg} and $\mu_{G|k}$ that is close in $\overline{W}_2$. To do this, first sample from the coupling of $\mu_{\mathrm{alg}}P, \mu_{G|k}$ to get $(\boldsymbol{X}', \boldsymbol{Y})$. Let P' be the time-reversal of P with respect to μ_{alg}. By Lemma 19, we can apply P' to $\boldsymbol{X}'$ to get a k^--independent set distributed as μ_{alg}, which we call $\boldsymbol{X}$. Again by Lemma 19, since P' only removes $O(\log\log n)$ nodes,

$$\begin{aligned}\mathop{\mathbf{E}}_{(\boldsymbol{X},\boldsymbol{Y})\sim\pi}\left\|\frac{\boldsymbol{X}}{m_2(\mu_{\mathrm{alg}})} - \frac{\boldsymbol{Y}}{m_2(\mu_{G|(-a,a)})}\right\|^2 &= 2 - 2\frac{\mathbf{E}_\pi\langle \boldsymbol{X}, \boldsymbol{Y}\rangle}{m_2(\mu_{\mathrm{alg}})m_2(\mu_{G|(-a,a)})} \\ &\leqslant 2 - 2\frac{(1-o(1))(\log n - O(\log\log n))}{(1-o(1))\log n} = o(1)\end{aligned}$$

because as long as $\boldsymbol{X}' = \boldsymbol{Y}$ in the coupling of $\mu_{\mathrm{alg}}P$ and $\mu_{G|k}$, after applying P' to $\boldsymbol{X}'$ to get $\boldsymbol{X}$, $\langle \boldsymbol{X}, \boldsymbol{Y}\rangle$ will be $\log n - O(\log\log n)$. Since the above argument is true for any $k \in k^* \pm a$, by Lemma 17, $\overline{W}_2(\mu_{\mathrm{alg}}, \mu_{G|(-a,a)})^2 = o(1)$. We will now show $\overline{W}_2(\mu_{G|(-a,a)}, \mu_G)^2 = o(1)$, and we will be done by the triangle inequality. First, observe that

$$m_2(\mu_G)^2 \in \log n \pm O(\log\log n) + O\left(2^{-a}\log n\right) = (1 \pm o(1))\log n.$$

because $\mathrm{d_{TV}}(\mu_G\,,\,\mu_{G|(-a,a)}) = O(2^{-a})$ and the largest independent set in $\boldsymbol{G}$ has size $O(\log n)$ with probability $1 - o(1)$ by Lemma 22. To conclude, using the coupling between $\mu_G, \mu_{G|(-a,a)}$ from the $O(2^{-a})$ TV distance, we see that

$$\overline{W}_2(\mu_G, \mu_{G|(-a,a)})^2 \leqslant 2 - 2\frac{(1 - O(2^{-a}))(\log n - O(\log\log n))}{(1 \pm o(1))\log n} = o(1).$$

■

Appendix E. Combinatorial Lemmas

In this section, we prove some combinatorial lemmas used in our concentration arguments above.

Lemma 26 *For any set of d size-ℓ subsets $A_1, \dots, A_d \in \binom{[m]}{\ell}$ with a large enough union $|A_1 \cup \cdots \cup A_d| > d\ell - d/2$, there exists $j \in [d]$ such that A_j is disjoint from the rest of the A_is.*

Proof We prove the contrapositive. Consider a graph H constructed by sequentially adding vertices $1 \dots, d$: when vertex i is added, we add the edge (i, j) for the smallest $j < i$ such that $A_i \cap A_j \neq \emptyset$ (if no such j exists we add no edge). To each such edge (i, j), we can assign at least one element of A_i which is a duplicate of an element already appearing in $A_1 \cup \cdots \cup A_{i-1}$. Hence $|A_1 \cup \cdots \cup A_j| \leqslant d - |E(H)|$. If no A_i is disjoint from all the rest, then the minimum degree in H is at least 1, and since $2|E(H)| = \sum_{i=1}^d \deg_H(i) \geqslant d$, we have our conclusion.

■

If A is a set, then we use the notation $\binom{A}{2} = \{\{a, b\} \mid a, b \in A,\ a \neq b\}$.

Lemma 27 *Let $A_1, \dots, A_d \in \binom{[m]}{\ell}$, and suppose that $|A_1 \cup \dots \cup A_d| = d\ell - a$, and that $d\ell < m$. Then*

$$\left|\binom{A_1}{2} \cup \dots \cup \binom{A_d}{2}\right| \geqslant d\binom{\ell}{2} - \min\left\{\frac{a}{\ell}\binom{\ell}{2}, \binom{a}{2}\right\}.$$

Proof Let $S = \emptyset$, and consider adding the sets A_i one at a time to S, starting with A_1, and ending with A_d. Let $B_i = A_i \setminus (A_1 \cup \dots \cup A_{i-1})$ be the new vertices added to S in step i, and call $|B_i| = \ell - a_i$. By definition,

$$|S| = \sum_{i=1}^{d} |B_i| = \sum_{i=1}^{d} \ell - a_i = d\ell - \sum_{i=1}^{d} a_i,$$

so $a = \sum a_i$.

We will now track the number of pairs added to $\bigcup_{i=1}^{d} \binom{A_i}{2}$ during this process. At step i, we add all pairs $\{a, b\} \in \binom{A_i}{2}$, excluding those that already appear in $\binom{A_i \setminus B_i}{2}$. Thus, the number of pairs added in step i is $\binom{\ell}{2} - \binom{a_i}{2}$. The result now follows because

$$\sum_{i=1}^{d} \binom{a_i}{2} \leqslant \min\left\{\frac{a}{\ell}\binom{\ell}{2}, \binom{a}{2}\right\}$$

where the last inequality follows because $a = \sum a_i$, $\binom{x}{2} + \binom{y}{2} \leqslant \binom{x+y}{2}$, and $a_i \leqslant \ell$ for all $i \in [d]$. ■

Lemma 28 *Let N_a be the number of ways to choose $A_1, \dots, A_d \in \binom{[m]}{\ell}$ such that $|A_1 \cup \dots \cup A_d| = d\ell - a$. Then*

$$N_a \leqslant \binom{m}{d\ell - a}\binom{d\ell - 1}{a - 1}\frac{(d\ell)!}{(\ell!)^d}.$$

Proof First we choose the elements in $T = A_1 \cup \dots \cup A_d$, for which there are $\binom{m}{d\ell - a}$ choices. Now we must choose the multiplicity with which each element in T appear in the multiset-union of the A_i; each element must have multiplicity at least 1, and then there are $\sum_{i=1}^{d} |A_i| - |T| = a$ excess appearances that we must partition among the $d\ell - a$ elements of T. By a stars-and-bars argument, the number of ways to distribute these excess appearances is at most $\binom{d\ell-1}{a-1}$. Finally, letting T' be the multiset resulting from T when each element is duplicated according to its multiplicity, the number of ways to form an ordered partition of T' into d subsets of size ℓ is at most $(d\ell)!/(\ell!)^d$. The product of these bounds gives our upper bound. ■

Appendix F. Algorithm lemmas

Lemma *[Restatement of Lemma 11]* *Let $\mathbf{G} \sim G(n, 1/2)$, $C > 6$. Then with probability $1 - o(1)$, all sets $S \subseteq [n]$ of size $0, \dots, \log n - C \log\log n$ satisfy $|\mathbf{deg}^{\uparrow}(S) - d_{|S|}| \leqslant d_{|S|}^{2/3}$.*

Proof Fix a set S of size j. Then observe that $\mathbf{deg}^{\uparrow}(S)$ is distributed as $\mathrm{Binom}(n - j, 2^{-j})$. So by a standard Chernoff bound,

$$\mathbf{Pr}[|\mathbf{deg}^{\uparrow}(S) - d_j| \geqslant \delta_j d_j] \leqslant 2\exp(-\delta_j^2 d_j/3).$$

Set $\delta_j = d_j^{-1/3}$. There are $\binom{n}{j} \leqslant 2^{j \log n}$ subsets of size j. The union bound tells us that

$$\mathbf{Pr}\left[\exists S \in [n]^{\leqslant \log n - C \log\log n} \text{ s.t. } \frac{|\mathbf{deg}^{\uparrow}(S) - d_{|S|}|}{d_{|S|}} \geqslant \delta_{|S|}\right] \leqslant \sum_{j=0}^{\log n - C\log\log n} \binom{n}{j} \exp(-\delta_j^2 d_j/3)$$

$$\leqslant \sum_{j=0}^{\log n - C \log\log n} \exp(\Theta(j \log n) - d_j^{1/3}/3).$$

Let the jth term be T_j. Observe that T_j is increasing because $j \log n$ is increasing in j, and d_j is decreasing in j. Thus it suffices to understand the last term. Let $k = \log n - C \log\log n$. Then

$$d_k = (n-k)2^{-k} = \frac{n-k}{n} \log^C n.$$

Moreover,

$$\Theta(k \log n) = \Theta(\log^2 n).$$

Therefore, if $C > 6$ is a constant, then $T_k = \exp(-\Omega(\log^{C/3-2} n)) = o(1)$, so we are done. ■