

Fast Mixing in Sparse Random Ising Models

Kuikui Liu^{*} Sidhanth Mohanty[†] Amit Rajaraman[‡] David X. Wu[§]

August 7, 2024

Abstract

Motivated by the community detection problem in Bayesian inference, as well as the recent explosion of interest in spin glasses from statistical physics, we study the classical Glauber dynamics for sampling from Ising models with sparse random interactions. It is now well-known that when the interaction matrix has spectral diameter less than 1, Glauber dynamics mixes in $O(n \log n)$ steps. Unfortunately, such criteria fail dramatically for interactions supported on arguably the most well-studied sparse random graph: the Erdős–Rényi random graph $G(n, d/n)$. There is a scarcity of positive results in this setting due to the presence of almost linearly many outlier eigenvalues of unbounded magnitude.

We prove that for the *Viana–Bray spin glass*, where the interactions are supported on $G(n, d/n)$ and randomly assigned $\pm\beta$, Glauber dynamics mixes in $n^{1+o(1)}$ time with high probability as long as $\beta \leq O(1/\sqrt{d})$, independent of n . We further extend our results to random graphs drawn according to the 2-community stochastic block model, as well as when the interactions are given by a “centered” version of the adjacency matrix. The latter setting is particularly relevant for the inference problem in community detection. Indeed, we build on this result to demonstrate that Glauber dynamics succeeds at recovering communities in the stochastic block model in a companion paper [LMR⁺24].

The primary technical ingredient in our proof is showing that with high probability, a sparse random graph can be decomposed into two parts — a *bulk* which behaves like a graph with bounded maximum degree and a well-behaved spectrum, and a *near-forest* with favorable pseudorandom properties. We then use this decomposition to design a localization procedure that interpolates to simpler Ising models supported only on the near-forest, and then execute a pathwise analysis to establish a modified log-Sobolev inequality.

^{*}MIT. Email: liukui@mit.edu

[†]MIT. Email: sidm@mit.edu. Supported by CSAIL.

[‡]MIT. Email: amit_r@mit.edu. Supported by an Akamai Presidential Fellowship.

[§]UC Berkeley. Email: david_wu@berkeley.edu. Supported by NSF Graduate Research Fellowship DGE-2146752.

Contents

1	Introduction	1
1.1	Diluted spin glasses	2
1.2	Centered graphs and community detection	3
1.3	Related work	4
2	Technical overview	5
3	Preliminaries	14
3.1	Notation	14
3.2	Markov chains	15
3.3	Ising models	17
3.4	Localization schemes	19
3.5	Graphs	20
4	Glauber dynamics on near-forest decompositions	23
4.1	On decomposing random graphs	23
4.2	Proving an MLSI for diluted spin glasses	25
4.3	Covariance bounds on near-forests	27
4.4	Entropy conservation for near-forest decompositions	32
4.5	MLSI for Ising models on pseudorandom near-forests	33
5	Glauber dynamics for centered stochastic block models	36
5.1	Technical overview for centered stochastic block models	37
5.2	A crash course on spectral independence	39
5.3	MLSI for Ising models on perturbed near-forests	40
5.4	Entropy conservation for centered stochastic block models	44
6	Near-forest decomposition for stochastic block models	46
6.1	Excisions of sparse random graphs are near-forests	48
6.2	Bounding the sizes of components	53
	References	56
A	Proofs omitted from the main text	60
A.1	Generalized MLSI for Ising models on pseudorandom near-forests	60
A.2	Bounds on various probabilities in sparse random graphs	62

1 Introduction

In this paper, we tackle the problem of algorithmically sampling from the Gibbs distribution of an *Ising model*, the quintessential model in statistical mechanics. These are probability distributions of the form

$$\mu_\beta(\sigma) = \mu_{\beta J, h}(\sigma) \propto \exp\left(\frac{\beta}{2}\sigma^\top J\sigma + \langle h, \sigma \rangle\right), \quad \forall \sigma \in \{\pm 1\}^n,$$

where $J \in \mathbb{R}^{n \times n}$ is a symmetric *interaction matrix*, $h \in \mathbb{R}^n$ is an *external field* vector, and $\beta \geq 0$ is the *inverse temperature*, which scales the interaction strengths. There is a vast literature devoted to their study due to their numerous intimate connections with physics, statistical inference, combinatorial optimization, and the study of constraint satisfaction problems in theoretical computer science; see e.g., [EKPS00, WJ08, MM09, Tal11, MNS15, FV17] and references therein. We consider the following fundamental question.

Question. Under what conditions on β , J , and h is (approximately) sampling from μ_β computationally tractable?

This question has a long and rich history, and there is now a good understanding in the worst-case setting. Sampling from μ_β is easy if at least one of the following conditions hold:

- J is entrywise nonnegative (and all entries of h have the same sign), i.e., the model is *ferromagnetic* [JS93].
- The interaction strengths are sufficiently weak, e.g., *Dobrushin's condition* $\beta \max_i \sum_j |J_{ij}| < 1$ is satisfied [Dob68, BD97].

Conversely, if one lets $J = -A_G$ where A_G is the adjacency matrix of a d -regular graph G , then there is a precise critical threshold $\beta_c(d) \approx 1/d$ such that the sampling problem is NP-hard when $\beta > \beta_c(d)$ via reduction from MAXCUT [SS12], which corresponds to the $\beta \rightarrow +\infty$ limit. Notably, nearly-linear time samplers exist when $\beta < \beta_c(d)$ [CLV21, AJK⁺21b], and so the problem can exhibit sharp *computational phase transitions*.

Motivated by the stochastic block model from the study of community detection, as well as the theory of spin glasses in statistical physics, we focus on the *average-case* setting where J is a random matrix and the goal is to successfully sample from μ_β with probability $1 - o_n(1)$ over the randomness of J . Furthermore, since our aim is to obtain fast samplers, we consider *Glauber dynamics* (or the *Gibbs sampler*), a natural Markov chain over $\{\pm 1\}^n$ used ubiquitously both in theory and in practice. Its evolution is described as follows: in each step, we select a uniformly random coordinate $i \in [n]$, and resample its assignment σ_i conditioned on all remaining coordinates σ_{-i} .

A recent line of work [BB19, EKZ22, CE22, KLR22, BBD23, AJK⁺24] has identified that having spectral diameter bounded as

$$\text{(Spectral Condition)} \quad \beta \cdot (\lambda_{\max}(J) - \lambda_{\min}(J)) < 1 \tag{1}$$

is sufficient to ensure that Glauber dynamics mixes in $O(n \log n)$ steps. More generally, polynomial-time samplers exist if βJ only has $O(1)$ many “outlier” eigenvalues exceeding 1; see [KLR22] for

a precise statement. These types of conditions are appealing since many classical random matrix ensembles (e.g., $\text{GOE}(n)$, the uniformly random d -regular graph, etc.) have tame bulk spectra. The spectral condition Eq. (1) is also sharp for rapid mixing of Glauber dynamics, as witnessed by the Curie–Weiss model where $J = \frac{1}{n}\mathbf{1}\mathbf{1}^\top$. Stronger evidence for hardness of sampling when Eq. (1) fails was recently given [Kun23].

1.1 Diluted spin glasses

Unfortunately, existing spectral criteria like Eq. (1) fail badly for interaction matrices supported on the edges of arguably the most well-studied sparse random graph: the Erdős–Rényi random graph $G \sim G(n, d/n)$ with bounded average degree $d \leq O(1)$. For instance, consider the setting where all entries of J are supported on a graph G , and randomly assigned ± 1 (or $\mathcal{N}(0, 1)$) independently. By varying the choice of G , this recovers several famous models of spin glasses in statistical physics, e.g., the Sherrington–Kirkpatrick model when $G = K_n$, and the Edwards–Anderson model when $G = \mathbb{Z}^d$. When $G \sim G(n, d/n)$, sometimes known as the Viana–Bray model (see e.g., [Mon98, MP01, MP03, GT04]), such matrices automatically have $\Omega(n^{1-o(1)})$ many eigenvalues of magnitude $\omega(1)$ with high probability due to an abundance of “stars” with unusually high degrees, despite the overall sparsity of G . Hence, a direct application of existing results would yield efficient samplers only when $\beta \leq o_n(1)$. For comparison, if G is a uniformly random d -regular graph, then Eq. (1) yields rapid mixing as long as $\beta \leq O(1/\sqrt{d})$, independent of n .

In this work, we overcome the abundance of large eigenvalues and obtain almost-linear mixing up to $\beta \leq O(1/\sqrt{d})$, thus resolving the discrepancy between $G(n, d/n)$ and the uniformly random d -regular graph. As above, our first main result concerns mixing of Glauber dynamics for a diluted spin glass given by randomly signing the edges of a sparse Erdős–Rényi graph.

Theorem 1.1. *Fix a constant $d > 1$. There exists a universal constant $C > 0$ such that for all $\beta \leq C/\sqrt{d}$ the following is true. Let A be the random matrix such that each off-diagonal entry is equal to 0 with probability $1 - d/n$, and is equally likely to be ± 1 with the remaining probability. With high probability over the draw of A , Glauber dynamics run on $\mu_{\beta A}$ mixes in time $n^{1+o_n(1)}$.*

The bound on the constant C we obtain is equal to $(3 - \sqrt{8})(1 - o_d(1)) \approx 0.17 - o_d(1)$. We expect that this bound can be improved using recent techniques for beating the spectral condition (1) for the Sherrington–Kirkpatrick model [AKV24]. Based on replica-symmetry breaking predictions from statistical physics, e.g., [GT04], we further conjecture that Glauber dynamics mixes in polynomial time if $d \tanh^2(\beta) < 1$, and takes exponential time to mix above this threshold.

Our results extend immediately to sparse random graphs drawn from the stochastic block model $\text{SBM}(n, d, \lambda)$; see Theorem 4.1. Roughly speaking, the key feature of such graphs which allows us to go beyond the spectral condition Eq. (1) is *sparsity* and *localization* of the eigenvectors corresponding to large eigenvalues.

Remark 1.2. We further believe that our methods extend to the regime where d is slightly super-constant, in particular when $d < \log n$. In the setting where $d > \log n$, the spectral condition (Eq. (1)) can be used to bound the mixing time for $\beta = O\left(\frac{1}{\sqrt{d}}\right)$.

1.2 Centered graphs and community detection

Besides the spin glass setting, we also consider the *centered* setting, which is of particular relevance to community detection. Let us recall the (2-community) *stochastic block model*, a heavily studied Bayesian model of community detection. We refer interested readers to [Abb18] for further details on the history and importance of the model.

- Let $\sigma \in \{\pm 1\}^n$ be a *signal* vector drawn uniformly at random, which intuitively encodes the ground truth community assignments. In the language of Bayesian inference, the uniform distribution over $\{\pm 1\}^n$ is the *prior*.
- Given σ , we draw a random graph G by including an edge between $u, v \in [n]$ independently with probability $\frac{d+\lambda\sqrt{d}}{n}$ if $\sigma(u) = \sigma(v)$, and with probability $\frac{d-\lambda\sqrt{d}}{n}$ otherwise, where $d, \lambda \in \mathbb{R}$ are fixed constants.

The goal of statistical inference is to recover the true community assignments σ given access to the noisy observation G . The entirety of this generative process induces a distribution over graphs denoted $\text{SBM}(n, d, \lambda)$, with the special case $\lambda = 0$ recovering the Erdős–Rényi graph $G(n, d/n)$. Note that having $\lambda \geq 0$ encourages the resulting random graph G to have more intracommunity edges than intercommunity edges. The magnitude of λ controls the *signal-to-noise ratio* (SNR), and is parameterized so that $\lambda^2 = 1$ corresponds to the famous *Kesten–Stigum threshold* (see e.g., [DKMZ11, MNS15]).

It is well-known that the inference problem can be reduced to the problem of sampling from the *posterior distribution* over $\{\pm 1\}^n$ given G , which for large d is approximately given by the Ising Gibbs measure $\mu(\sigma) \propto \exp\left(\frac{\beta}{2}\sigma^\top \left(A_G - \frac{d}{n}\mathbf{1}\mathbf{1}^\top\right)\sigma\right)$, where $\beta = \beta(d, \lambda)$ is some explicit function monotone increasing in λ .¹ A natural and longstanding open problem in the field is giving provable guarantees for recovering σ by simulating Glauber dynamics. One of the principal difficulties is that the regime $\lambda > 1$, where nontrivial recovery is information-theoretically possible, is also well within the regime in which worst-case mixing of Glauber dynamics is exponential in n . Hence, the vast majority of existing results in the literature give guarantees for algorithms not based on Markov chains at all (such as belief propagation and spectral algorithms, see, e.g., [DKMZ11, Mas14, MNS14, AS15, BLM15, MNS15, MNS18, YP23] and references therein).

Towards understanding the performance of Glauber dynamics for inference, our second main contribution concerns sampling from the Ising model corresponding to the centered adjacency matrix $J = A_G - \mathbb{E}[A_G|\sigma]$.² In a companion paper [LMR⁺24], we leverage this mixing result to prove that with high probability over G , a polynomial number of steps of Glauber dynamics for a rescaled version of the true posterior distribution achieves $\Omega(1)$ correlation with the ground truth σ (or $-\sigma$) as long as λ exceeds a certain (reasonable) constant greater than 1.

¹Technically, it is given by $\beta(d, \lambda, n) = \frac{1}{2} \ln\left(\frac{d+\lambda\sqrt{d}}{d-\lambda\sqrt{d}}\right) - \frac{1}{2} \ln\left(\frac{n-d+\lambda\sqrt{d}}{n-d-\lambda\sqrt{d}}\right)$, which tends to $\beta(d, \lambda) = \frac{1}{2} \ln\left(\frac{d+\lambda\sqrt{d}}{d-\lambda\sqrt{d}}\right) \approx \frac{\lambda}{\sqrt{d}}$ at an $O(1/n)$ rate as $n \rightarrow \infty$ for fixed d, λ .

²Note that (up to a diagonal shift), $\mathbb{E}[A_G|\sigma] = \frac{d}{n}\mathbf{1}\mathbf{1}^\top + \frac{\lambda\sqrt{d}}{n}\sigma\sigma^\top$ and $\mathbb{E}[A_G] = \frac{d}{n}\mathbf{1}\mathbf{1}^\top$, so the former contains an uninformative eigenspace parallel to $\mathbf{1}$, as well as the signal σ .

Theorem 1.3. Fix constants $d > 1, \lambda \geq 0$. There exists a universal constant $C > 0$ such that for all $\beta \leq C/\sqrt{d}$ the following is true. Let $\bar{A}_G := A_G - \mathbf{E}[A_G|\sigma]$ be the centered adjacency matrix of $G \sim \text{SBM}(n, d, \lambda)$. With high probability over the draw of G , Glauber dynamics run on $\mu_{\beta \bar{A}_G}$ mixes in time $n^{1+o_d(1)}$.

1.3 Related work

We now survey some of the literature relevant to the study of Ising models and sampling via Markov chains.

Sampling from spin systems on sparse random graphs. In the ferromagnetic case where $J = A_G$, Mossel & Sly [MS13] established a mixing time of $n^{1+\Theta(1/\log \log n)}$ for Glauber dynamics in the regime $d \tanh(\beta) < 1$ with high probability over $G \sim G(n, d/n)$. This is sharp as the mixing time necessarily scales exponentially in n when $d \tanh(\beta) > 1$. More recently, when J is supported on $G \sim G(n, d/n)$ and has i.i.d. Gaussian entries, Efthymiou & Zampetakis [EZ24] established almost-quadratic mixing of Glauber dynamics when $\beta \leq \sqrt{2\pi}/d$, and conjectured that their regime of β is optimal for rapid mixing. While Theorem 1.1 as stated only considers random signs, we believe that an adaptation of our techniques refutes this conjecture. For related results on the hardcore gas model (i.e., random independent sets) on $G(n, d/n)$, see [EF23] and references therein.

In a recent work, Bauerschmidt, Bodineau & Dagallier [BBD23] proved that the Kawasaki dynamics for the ferromagnetic Ising model with fixed magnetisation on a random d -regular graph mixes rapidly for $\beta < O(1/\sqrt{d})$.

Stochastic localization and sampling. A common strategy to analyze a Markov chain is to decompose the stationary distribution as a mixture of simpler distributions. To this end, we employ the *stochastic localization* process. This idea was first introduced in the context of studying isoperimetry in convex bodies by Eldan [Eld13], and has since seen numerous applications. The work of Eldan, Koehler & Zeitouni [EKZ22] introduced it to the study of Glauber dynamics for Ising models, and used it to prove that any Ising model can be expressed as a “nice” mixture of rank-1 Ising models, a fact that was key to establishing the spectral condition for rapid mixing of Glauber dynamics. Later, Chen & Eldan [CE22] provided a different proof of the same result by giving a stochastic localization process to decompose an Ising model into a mixture of product distributions; see also [BB19], which proved a similar result using a similar decomposition.

A parallel line of work uses stochastic localization as an algorithmic technique for sampling from spin glass models, rather than as a tool for analyzing Markov chains. This line began with the work of El Alaoui, Montanari & Sellke [EAMS22] on an algorithm to sample from a distribution close to the Sherrington–Kirkpatrick model (SK model) in Wasserstein distance, when $\beta < 1$, based on discretizing the stochastic localization process. In contrast, the spectral condition implies that Glauber dynamics succeeds only when $\beta < 1/4$, albeit under the more stringent total variation distance. Analogous results have also been obtained for using stochastic localization to sample from mixed p -spin Ising models [EAMS23], parallel sampling algorithms [AHL⁺23], posteriors of spiked matrix models [MW23], and spherical spin glasses [HMP24]. See [Mon23] for a

survey on recent developments, and connections to diffusion models in machine learning.

2 Technical overview

We now describe how we prove [Theorem 1.1](#) on the rapid mixing of Glauber dynamics for an Ising model on a randomly signed sparse Erdős–Rényi graph. The proof of [Theorem 1.3](#) for centered stochastic block model graphs, while more involved, follows similar inspiration.

Modified log-Sobolev inequality. The rapid mixing of Glauber dynamics follows from a *modified log-Sobolev inequality* (MLSI), which is a functional inequality relating the global behavior of any function to its “local differences” in the Markov chain; see, e.g., [BT06, Corollary 2.8] for the connection between MLSI and mixing times. Concretely, for a function $f : \{\pm 1\}^n \rightarrow \mathbb{R}_{>0}$ and probability distribution μ , the quantities of interest to us are its:

- **Local entropy.** The *local entropy* of f is $\mathcal{E}_\mu(f, \log f) := \mathbf{E}_{x,y} \left[(f(x) - f(y)) \cdot \log \frac{f(x)}{f(y)} \right]$ where $x \sim \mu$ and y is sampled by taking a single step of the Glauber dynamics chain from x .
- **Global entropy.** The *global entropy* of f is $\text{Ent}_\mu[f] := \mathbf{E}_\mu f \log f - \mathbf{E}_\mu f \log \mathbf{E}_\mu f$.

We say that Glauber dynamics for μ satisfies a *modified log-Sobolev inequality* (MLSI) if for any function $f : \{\pm 1\}^n \rightarrow \mathbb{R}_{>0}$:

$$\mathcal{E}_\mu(f, \log f) \gtrsim \text{Ent}_\mu[f],$$

where the \gtrsim hides a $\text{poly}(n)$ factor.

The following is the functional inequality we prove, which also implies rapid mixing.

Theorem 2.1 (Informal version of [Theorem 4.1](#)). *Let A be the adjacency matrix of a randomly signed graph $G \sim \mathcal{G}(n, d/n)$. There exists a constant $\beta = \Omega(1/\sqrt{d})$ such that $\mu = \mu_{\beta A}$ satisfies an MLSI.*

Remark 2.2. In fact, we show that the polynomial factor hidden by \gtrsim in the above is $n^{-1-o(1)}$. However, we do not stress this in the overview for the sake of simplicity.

The general strategy to prove this inequality in our setting is similar to that of [EKZ22, AJK⁺21a, CE22], which prove similar inequalities for Glauber dynamics on other Ising models, and can be summarized by the following.

Decompose μ into a mixture ρ over “simpler” distributions μ_z where $\mu = \mathbf{E}_{z \sim \rho} \mu_z$, and then establish the following chain of inequalities:

$$\mathcal{E}_\mu(f, \log f) \underset{\substack{\uparrow \\ \text{Conservation} \\ \text{of local} \\ \text{entropy}}}{\geq} \mathbf{E}_{z \sim \rho} \mathcal{E}_{\mu_z}(f, \log f) \underset{\substack{\uparrow \\ \text{MLSI for} \\ \text{simple} \\ \text{distributions}}}{\gtrsim} \mathbf{E}_{z \sim \rho} \text{Ent}_{\mu_z}[f] \underset{\substack{\uparrow \\ \text{Conservation} \\ \text{of entropy}}}{\gtrsim} \text{Ent}_\mu[f].$$

The first inequality is generic and does not use any properties of the structure of ρ ; see, e.g., [AJK⁺21a, Page 19]. We reproduce the details here for completeness:

$$\mathcal{E}_\mu(f, \log f) = \sum_{x \sim y} \frac{1}{n} \cdot \frac{\mu(x)\mu(y)}{\mu(x) + \mu(y)} \cdot (f(x) - f(y)) \cdot \log \frac{f(x)}{f(y)}$$

$$\begin{aligned}
&= \sum_{x \sim y} \frac{1}{n} \cdot \frac{\mathbf{E}_{z \sim \rho} [\mu_z(x)] \cdot \mathbf{E}_{z \sim \rho} [\mu_z(y)]}{\mathbf{E}_{z \sim \rho} [\mu_z(x)] + \mathbf{E}_{z \sim \rho} [\mu_z(y)]} \cdot (f(x) - f(y)) \cdot \log \frac{f(x)}{f(y)} \\
&\geq \sum_{x \sim y} \frac{1}{n} \cdot \mathbf{E}_{z \sim \rho} \left[\frac{\mu_z(x) \mu_z(y)}{\mu_z(x) + \mu_z(y)} \right] \cdot (f(x) - f(y)) \cdot \log \frac{f(x)}{f(y)} \\
&= \mathbf{E}_{z \sim \rho} [\mathcal{E}_{\mu_z}(f, \log f)],
\end{aligned}$$

where the inequality follows from concavity of the function $(a, b) \mapsto \frac{ab}{a+b}$ in the nonnegative quadrant.

The second inequality relies on the MLSI for the component measures, which we hope are easier to analyze. The third inequality necessarily relies on the properties of ρ . We now describe our construction of a measure decomposition of μ when it is an Ising model with PSD interaction matrix J and external field h , and then show how to establish the MLSI for the simple distributions, and conservation of entropy.³

Measure decomposition. Our measure decomposition is based on the *Hubbard–Stratonovich transform*, an algebraic trick to express a Gibbs distribution arising from a quadratic Hamiltonian as a mixture of simpler distributions. The measure decomposition takes the following form.

Let M be a PSD matrix that we shall call a *control matrix*, let $x \sim \mu$, and let $z := x + M^{-1/2}g$ for $g \sim \mathcal{N}(0, I)$ independent of x . Let ρ denote the distribution of z and let μ_z denote the distribution of $x|z$.

We now perform a fairly mechanical calculation to determine the distribution of $x|z$, just to illustrate a favorable algebraic cancellation. Let $P(x)$, $P(z|x)$ and $P(x, z)$ be the joint densities of x , $z|x$ and (x, z) respectively. We have:

$$\begin{aligned}
\Pr[x = x|z = z] &\propto P(x, z) \\
&= P(x) \cdot P(z|x) \\
&\propto \exp\left(\frac{1}{2}x^\top Jx + \langle h, x \rangle\right) \cdot \exp\left(-\frac{1}{2}(z - x)^\top M(z - x)\right) \\
&\propto \exp\left(\frac{1}{2}x^\top (J - M)x + \langle h + Mz, x \rangle\right).
\end{aligned}$$

Designing the control matrix The upshot of the above is that the measures μ_z are Ising models with interaction matrix $J - M$ and some external field depending on z . We would like to design M such that

- It is relatively easy to prove an MLSI for an Ising model with interaction matrix $J - M$.
- The mixture ρ satisfies conservation of entropy.

³We can assume that $J \succeq 0$ without loss of generality, as adding an arbitrary diagonal matrix to J does not change the Gibbs distribution.

Choosing $M = J$ makes the interactions disappear altogether, and makes μ_z a product distribution. This is desirable as every product distribution satisfies an MLSI. Indeed, this choice is made in the work of Chen and Eldan [CE22, Corollary 51] where they prove an MLSI for Glauber dynamics on Ising models when J has spectral diameter at most 1 (see Eq. (1)). They then use the bound on the spectral diameter to prove conservation of entropy for the resulting product decomposition of the Ising model.

Unfortunately, when J is a matrix associated with a sparse Erdős–Rényi graph, the abnormally high-degree vertices cause a large number of outlier eigenvalues, which is a hurdle for their technique to establish conservation of entropy. Nevertheless, studying their technique motivates our design of M .

Conservation of entropy. Recall that *conservation of entropy* for our decomposition refers to the following functional inequality. For $f : \{\pm 1\}^n \rightarrow \mathbb{R}_{>0}$:

$$\mathbf{E}_{z \sim \rho} \text{Ent}_{\mu_z}[f] \gtrsim \text{Ent}_{\mu}[f].$$

The strategy for doing this, as carried out in [EKZ22, CE22], is to chart a continuous path between μ and μ_z and control the derivative of entropy along this path. Concretely, there is a stochastic process $(\mu_t)_{0 \leq t \leq 1}$ on the space of probability distributions called *stochastic localization* with the following properties; see [CE22, Proposition 39 and Lemma 40]:⁴

- $\mu_0 = \mu$ and $\mu_1 = \mu_z$ for $z \sim \rho$.
- The function $F(t) = \mathbf{E} \text{Ent}_{\mu_t}[f]$, where the expectation is taken over the randomness in the choice of μ_t , satisfies the following lower bound on its derivative:

$$\left. \frac{dF}{dt} \right|_t \geq - \sup_h \left\| M^{1/2} \cdot \mathbf{Cov}(\mu_{J-tM,h}) \cdot M^{1/2} \right\|_{\text{op}} \cdot F(t). \quad (2)$$

If we can show for any $h \in \mathbb{R}^n$ and $t \in [0, 1]$ that $\|M^{1/2} \cdot \mathbf{Cov}(\mu_{J-tM,h}) \cdot M^{1/2}\|_{\text{op}} \leq q$ for some $q > 0$, then

$$F(1) \geq F(0) \cdot \exp(-q),$$

which, in more familiar terms, is

$$\mathbf{E}_{z \sim \rho} \text{Ent}_{\mu_z}[f] \geq \exp(-q) \text{Ent}_{\mu}[f].$$

Our goal thus reduces to designing M such that $\|M^{1/2} \cdot \mathbf{Cov}(\mu_{J-tM,h}) \cdot M^{1/2}\|_{\text{op}}$ is small, and that it is easy to prove an MLSI for Ising models with interaction matrix $J - M$.

Intuitively, the large spectral norm of M itself can be blamed on the high-degree vertices. To heuristically get a sense of where $\mathbf{Cov}(\mu_{J-tM,h})$ might have large spectral norm, let us pretend J is the interaction matrix of a tree, and consider the special case of $h = 0$ and $t = 0$. We remark that the final proof will *uniformly* bound the covariance matrix for all external fields $h \in \mathbb{R}^n$ and times $t \in [0, 1]$ with high probability over the graph, although we omit the details here for the sake of simplicity.

⁴ We refer the reader to [EAM22] for a nice exposition on stochastic localization and why it indeed charts such a path.

Under this pretense, via standard formulas for correlations in tree graphical models,

$$\mathbf{Cov}(\mu_{J,0})_{i,j} \approx \Omega\left(\frac{1}{\sqrt{d}}\right)^{\text{dist}(i,j)}.$$

Let $n_\ell(i)$ denote the number of neighbors of vertex i at distance exactly ℓ from i . The entries of the i th row corresponding to the distance- ℓ neighbors of a vertex i are at least $(1/d)^{\ell/2}$, and since the operator norm of a matrix is at least the ℓ_2 norm of any row, we have:

$$\|\mathbf{Cov}(\mu_{J,0})\|_{\text{op}} \gtrsim \Omega\left(\sqrt{\frac{n_\ell(i)}{d^\ell}}\right).$$

This can be huge if vertex i is abnormally high-degree, or more generally, if vertex i has an abnormally large ℓ -neighborhood. In particular, if $n_\ell(i) > (Cd)^\ell$ for some appropriately chosen constant $C > 1$, that would reflect in a blow-up in the spectral norm of $\mathbf{Cov}(\mu_{J,0})_{N_\ell(i)}$, the covariance matrix restricted to the ℓ -neighborhood of i . The idea then is to consider M that is supported only on rows and columns away from such ‘‘rogue’’ vertices i .

In particular, we partition the graph into two pieces, one piece with vertices close to the rogue vertices, and another piece with vertices far away from rogue vertices, and design M to be supported on the rows and columns from the second piece.

Graph decomposition. Concretely, for a graph G :

- For each vertex v , define $\ell(v)$ as the smallest value of ℓ such that for all $L \geq \ell$, the number of vertices at distance $\leq L$ from v is at most $(d(1 + \varepsilon))^L$.
- Set H as $\bigcup_{v \in V} B_{\ell(v)}(v)$, where $B_{\ell(v)}(v)$ is the radius $\ell(v)$ ball around v .

The above gives a natural partition of the graph G into $B := G \setminus H$, which we call the *bulk*, and H , which we call the *near-forest*. The point of this decomposition is that all the rogue vertices reside in the near-forest, and the vertices touched by edges in the bulk are ‘‘tame’’; see [Figure 1](#) for an illustration. We prove the following in the context of sparse random graphs.

Lemma 2.3. *For $G \sim G(n, d/n)$ and random signing c of the edges of G , there exists with high probability a subgraph H and $\varepsilon = O(\log d/d)^{1/3}$ such that:*

- Every connected component of H is a tree plus at most one additional edge.
- The diameter of each connected component of H is at most $O\left(\frac{\log n}{\varepsilon^3 d}\right)$.
- For any vertex v and $\ell \geq 1$, the size of $B_\ell(v)$ is at most $\Delta \cdot (d(1 + \varepsilon))^{\ell-1}$ for $\Delta = o(\log n)$.
- For any vertex v and $\ell \geq 1$, the size of $B_\ell(v) \cap B$ is at most $(d(1 + \varepsilon))^\ell$ where B is the set of all vertices touched by an edge in $G \setminus H$.
- The spectral radius of the signed adjacency matrix of $G \setminus H$ is at most $2\sqrt{d}(1 + \varepsilon)$.

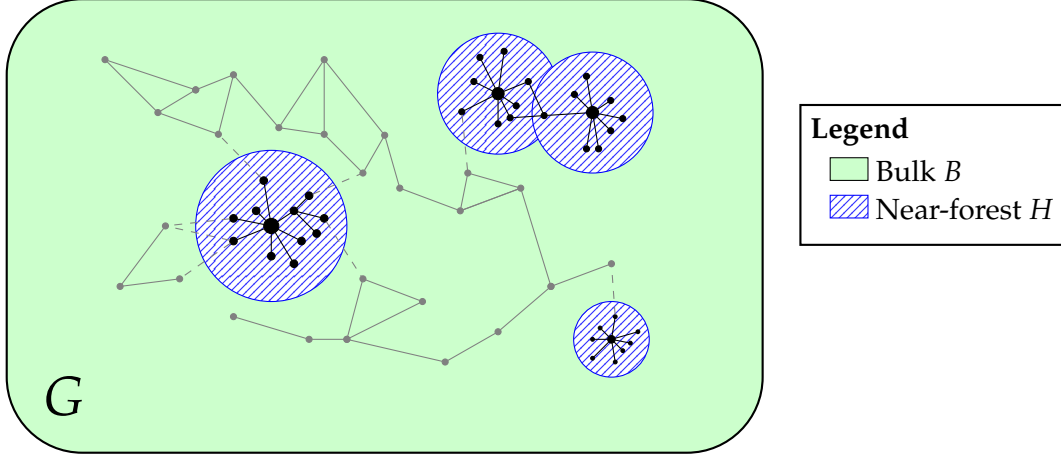


Figure 1: An illustration of our graph decomposition (Lemma 2.3) for a sparse SBM with average degree $d = 3$ and $\varepsilon = 0.5$. The bulk B (green) consists solely of vertices with degree at most $d(1 + \varepsilon) = 4.5$. The near-forest H (shaded blue circles) comprise the local neighborhoods of high-degree vertices (enlarged for emphasis) in G . The local neighborhoods of nearby high-degree vertices can merge, as shown by the two blue circles in the top right. Nevertheless, with high probability, all the connected components in H are trees with at most one additional edge.

We defer a discussion of the proof of this graph decomposition lemma to later in this overview, and focus on the proof of mixing of Glauber dynamics for now.

The above statement motivates the choice of the control matrix as:

$$M := \beta \cdot \left(A_{G \setminus H} + 2\sqrt{d}(1 + \varepsilon) \cdot I_{G \setminus H} \right).$$

An outline of the proof is presented in Figure 2. We remark that other proofs of MLSIs that utilize stochastic localization, such as that of (1), also follow a similar outline, albeit with many steps being much simpler. We shall now concretely clarify our goals before proceeding.

- Recall that our goal in proving conservation of entropy was to give an upper bound on

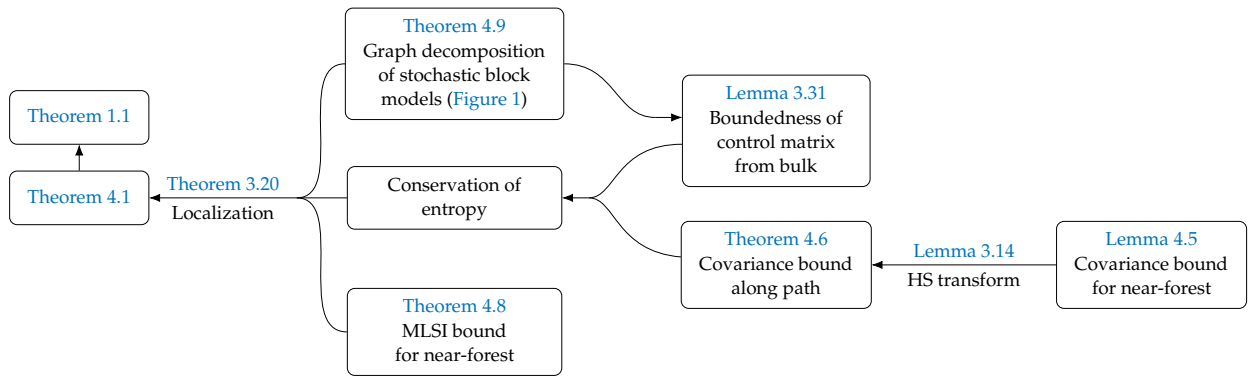


Figure 2: A flow chart outlining our proof of an MLSI for diluted spin glasses.

$$\|M^{1/2} \cdot \mathbf{Cov}(\mu_{J-tM,h}) \cdot M^{1/2}\|_{\text{op}}.$$

- We would also like to prove that any Ising model with interaction matrix $J - M$ satisfies an MLSI. In light of [Lemma 2.3](#), the structure of $J - M$ is a “near-forest” where every connected component is a low-diameter tree with at most one additional edge with “not too large” local neighborhoods. Our goal is to prove an MLSI for such near-forest Ising models.

Covariance bound along localization path. Observe that $\|M^{1/2} \cdot \mathbf{Cov}(\mu_{J-tM,h}) \cdot M^{1/2}\|_{\text{op}}$ is at most $\|\mathbf{Cov}(\mu_{J-tM,h}) \cdot M\|_{\text{op}}$. We show that this spectral norm is $O(1)$.

For simplicity, in this overview, we focus on the special case when $t = 0$ and $h = 0$. Let us denote $\mathbf{Cov}(\mu_{J,0})$ as Q . First observe that:

$$\|Q \cdot M\|_{\text{op}} \leq \|Q\|_{\text{op}} \cdot \|M\|_{\text{op}}.$$

To obtain a handle on Q , we perform a trick identical to how we constructed our measure decomposition, albeit for a completely different purpose now.

Let $\mathbf{x} \sim \mu_{J,0}$, and let $\mathbf{z} := \mathbf{x} + M^{-1/2}\mathbf{g}$ for $\mathbf{g} \sim \mathcal{N}(0, I)$.⁵

We obtain a bound on Q by using $\mathbf{Cov}[\mathbf{z}] \succeq \mathbf{Cov}[\mathbf{x}]$, which lets us pass to giving a bound on $\mathbf{Cov}[\mathbf{z}]$. To this end, we prove that the distribution of \mathbf{z} is *strongly log-concave* and give quantitative lower bound on its strong log-concavity. A bound on the covariance then immediately follows from the Brascamp–Lieb inequality, which says:

Suppose $\nabla^2 \log f(\mathbf{y}) \succeq \Gamma$ for all \mathbf{y} , where f is the density of \mathbf{z} . Then, $\mathbf{Cov}[\mathbf{z}] \preceq \Gamma^{-1}$.

A standard calculation reveals that it suffices to give a lower bound in the PSD order on the matrix $M - M \cdot \mathbf{Cov}(\mu_{J-M,0}) \cdot M$. Since M is supported only on the “bulk” $B := \mathbf{G} \setminus H$, we can write this as $M - M \cdot \mathbf{Cov}(\mu_{J-M,0})_B \cdot M$, where $\mathbf{Cov}(\mu_{J-M,0})_B$ is the principal submatrix of the covariance matrix on the rows and columns corresponding to non-isolated vertices in B .⁶ We now make the observation

$$M - M \cdot \mathbf{Cov}(\mu_{J-M,0}) \cdot M = M - M \cdot \mathbf{Cov}(\mu_{J-M,0})_B \cdot M \succeq M - M^2 \cdot \|\mathbf{Cov}(\mu_{J-M,0})_B\|_{\text{op}},$$

which motivates obtaining a bound on $\mathbf{Cov}(\mu_{J-M,0})_B$. The spectral norm of M can be suitably bounded using existing results on the spectra of sparse random graphs by a small enough constant that scales linearly with β ; see [Section 3.5](#) for details. We will show that $\|\mathbf{Cov}(\mu_{J-M,0})_B\|_{\text{op}} \leq O(1)$.

The Ising model on near-forests. We now discuss how to control the spectral behavior of $\mathbf{Cov}(\mu_{H,h})$ where H is a near-forest satisfying the conditions from [Lemma 2.3](#), as well as how we prove the MLSI for such Ising models.

⁵Technically, M needs to be slightly modified by adding a tiny diagonal on the entries outside $V(B)$, but we ignore that consideration here.

⁶We need the lower bound Γ to be strictly positive definite, which motivates the trick from the previous footnote of adding a small scaling of the identity matrix to the $V(H)$ block. However, we ignore this technicality for the present discussion.

Observe that $\mathbf{Cov}(\mu_{H,h})$ is a block diagonal matrix where each block corresponds to the covariance matrix of the Ising model of a connected component F , which is a tree with at most one excess edge. Each F has a specially identified boundary ∂F , which is the intersection of F with B . Thus, it is sufficient to bound the spectral norm of each $\mathbf{Cov}(\mu_{F,h})_{\partial F}$. Bounding the spectral norm of this matrix entails the following steps:

- We reduce to proving a covariance norm bound to the case when F is a tree, the interactions are nonnegative, and the external field is 0, but omit the description of this reduction in the overview; see the proof of [Lemma 4.5](#) for details.
- There is a simple explicit formula for the covariance matrix of an Ising model on a tree with zero external field in terms of the interaction matrix: for a pair of vertices i and j , and the unique path P between i and j , the covariance $\mathbf{Cov}(\mu_{F,0})_{i,j}$ is equal to $\prod_{e \in P} \tanh(J_e)$.
- We use the trace moment method to bound the desired spectral norm. The explicit formula from the previous point makes such a calculation tractable. When performing this calculation, one needs to count closed walks in the tree, and the bound on the size of balls in H guaranteed by [Lemma 2.3](#) lead us to the desired bound.

We now turn our attention to proving that the leftover near-forest in $J - M$ satisfies an MLSI for any choice of external field h . In order to make the discussion smoother and concretely point out which models we show an MLSI for, we make the following definition (see [Definition 4.3](#) for a formal version).

Definition 2.4 ((Δ, D) -pseudorandomness, informal). We say a graph H is (Δ, D) -pseudorandom if for every vertex $v \in V(H)$, the number of vertices at distance at most ℓ from v is at most $\Delta \cdot D^{\ell-1}$.

It suffices for us to prove an MLSI for (Δ, D) -pseudorandom near-forests where $\Delta = o(\log n)$ and $D = d(1 + \varepsilon)$. To show an MLSI for pseudorandom near-forests, we reduce to exhibiting an MLSI for pseudorandom trees; see [Section 4.5](#) for details.

We prove the following for pseudorandom trees.

Theorem 2.5 (Informal version of [Theorem 4.8](#)). Suppose $\Delta = o(\log n)$ and $0 \leq \gamma < 1$ is an absolute constant. Then, for any (Δ, D) -pseudorandom tree H with interaction matrix J with all entries in $\left(-\frac{\gamma}{\sqrt{D}}, \frac{\gamma}{\sqrt{D}}\right)$ and any external field h , Glauber dynamics for $\mu_{J,h}$ satisfies an MLSI.

To prove the above, we employ the strategy of decomposing the measure into a mixture of product distributions using stochastic localization. Recall from [\(2\)](#) that to control the derivative of the entropy along the stochastic localization path, it suffices to control $\|M \cdot \mathbf{Cov}(\mu_{J-tM,h})\|_{\text{op}}$ where M is the control matrix of choice for stochastic localization. It is tempting to use $J + \lambda_{\min}(J) \cdot I$ as the control matrix. However, the exact choice of the control matrix requires care because the parts of both this matrix and the covariance matrix that have high spectrum are on the same coordinates. In fact, we need to engineer the control matrix so that its high-spectral norm part partially cancels the corresponding part of the covariance matrix when they are multiplied. We refer the reader to the proof of [Lemma 4.19](#) for details. We spend the rest of the overview discussing how we prove [Lemma 2.3](#).

Proving the existence of a graph decomposition for sparse random graphs. We next outline how to prove [Lemma 2.3](#), which establishes nice properties of the graph decomposition into bulk and near-forest components for sparse stochastic block models. Before diving into the discussion, we restate the lemma for convenience.

Lemma 2.3. *For $G \sim G(n, d/n)$ and random signing c of the edges of G , there exists with high probability a subgraph H and $\varepsilon = O(\log d/d)^{1/3}$ such that:*

- *Every connected component of H is a tree plus at most one additional edge.*
- *The diameter of each connected component of H is at most $O\left(\frac{\log n}{\varepsilon^3 d}\right)$.*
- *For any vertex v and $\ell \geq 1$, the size of $B_\ell(v)$ is at most $\Delta \cdot (d(1 + \varepsilon))^{\ell-1}$ for $\Delta = o(\log n)$.*
- *For any vertex v and $\ell \geq 1$, the size of $B_\ell(v) \cap B$ is at most $(d(1 + \varepsilon))^\ell$ where B is the set of all vertices touched by an edge in $G \setminus H$.*
- *The spectral radius of the signed adjacency matrix of $G \setminus H$ is at most $2\sqrt{d}(1 + \varepsilon)$.*

As its name suggests, the bulk component has a tame spectrum, because every vertex in the bulk has bounded degree by construction. On the other hand, because sparse random graphs are locally tree-like, the near-forest component H consists of the union of balls around the high degree vertices. Let us see how to push this further to prove the other two properties of the graph decomposition.

To show that H is a near-forest, a crucial ingredient is the following structural property about small sets in sparse random graphs from [[BLM15](#), Lemma 30].

For $G \sim \text{SBM}(n, d, \lambda)$, with probability $1 - o(1)$, any connected set of vertices with fewer than $\frac{1}{2} \cdot \log_d n$ vertices has at most one cycle.

In fact, the above lemma can be used to rule out the existence of multiple cycles in any connected set of vertices (in particular the components of H), as long as the diameter of this set is $O(\log n)$. The idea is to rule out the existence of the second cycle by considering the small set of vertices consisting of the two cycles, along with the shortest path between them. Because of the diameter bound, the total size of this set is also $O(\log n)$, so the above lemma applies — if the hidden constant in the diameter bound is small enough, then the lemma implies that this subgraph has at most one cycle, a contradiction.

To summarize, we have boiled down the proof of [Lemma 2.3](#) to showing the diameter bound.

Bounding the diameter of the near-forest. We illustrate the intuition behind such a statement from the perspective of branching processes. Consider the following exploration process, which simulates the ball-growing process that generates H .

At all stages of the exploration, we maintain a set of active vertices, unexplored vertices, and inactive vertices. The active vertices represent the boundary of the connected component we are currently exploring.

1. If there are no active vertices, select a vertex v which has not been explored yet. If all vertices have been visited, terminate the process.
2. Select a vertex v in the active set, and iteratively query its radius- r neighborhoods.
3. If there is some vertex u at distance at most r which is r -heavy, that is, $|B_r(u)| > (d(1 + \varepsilon))^r$, then expand out the active vertices to include u by growing a ball of radius $2 \cdot \ell_u$ around v . Set all vertices on the interior of the active set to be inactive.
4. If no active vertices remain, this means that a new connected component of H has been fully generated.

Analyzing the exploration process directly is challenging, but one natural approach is to consider a branching process which simulates the above exploration process.

For simplicity, let us restrict to the case where G is a sparse random graph with constant average degree d . It is well known that for such a graph, the degree of any vertex v has Poisson tails, in that for any $s \geq 1$, we have

$$\Pr[\deg(v) \geq sd] \leq \exp(-\Omega(sd \log s)).$$

In particular, the maximum degree of G is $\Theta\left(\frac{\log n}{\log \log n}\right)$ with high probability. In fact, using standard techniques from the theory of branching processes, we prove a similar bound ([Lemma A.1](#)) for the sizes of radius- r neighborhoods of v :

$$\Pr[|B_r(v)| \geq (sd)^r] \leq \exp(-\Omega(ds^r)).$$

Observe how the above tail bound decays doubly exponentially in r . Using this, it is not too difficult to show that the probability that a vertex is r -heavy also decays as $\exp(-\Omega(ds^r))$. We also remark that the proof strategy of this tail bound is key in proving that with high probability, the size of *any* ball of radius ℓ is at most $o(\log n) \cdot (d(1 + \varepsilon))^{\ell-1}$, which is required to attain an $n^{1+o(1)}$ mixing time.

The hope in a proof involving branching processes would be that although the number of vertices at depth at most r is $O(d^r)$, the probability of any of these vertices being r -heavy decays doubly exponentially with r , so the radius of the largest ball that captures the vertex under consideration is very small — it has tail bounds that decay doubly exponentially, and has expectation $e^{-\Omega(d)}$. Consequently, the number of active vertices is likely to shrink in each step of the exploration. Furthermore, the $O(\log n/d)$ diameter bound is tight using such a proof technique: a vertex has slightly abnormal degree and is 1-heavy with probability $e^{-\Theta(d)}$, so with probability $1/n$, there are $O(\log n/d)$ such vertices that are all adjacent to each other.

However, this proof technique as stated is difficult to carry out, because the definition of heaviness involves not just subtrees rooted at the relevant vertex in the tree, but also subtrees rooted at ancestors and siblings, which become difficult to handle. For this reason, the proof we give in the body of the paper involves a more direct moment method for the diameter bound, but the intuition remains the same. We now briefly describe the ideas involved therein.

In our actual proof, we consider the associated *cluster graph*, which has as its vertex set those vertices v that have $\ell(v) > 0$, and edges between vertices if their corresponding balls intersect. Our goal then is to bound the diameter of this new cluster graph, where each edge is assigned weight equal to the sum of the radii of (the balls around) its endpoints. To do so, we consider the probability that a *fixed* collection of vertices forms a long path. The key is that if these balls have small radii, then they must intersect in a specific fashion. However, each of the balls are essentially random (small!) subsets of the n vertices, so this is extraordinarily unlikely. On the other hand, if the radii are somewhat large, the doubly exponential tail bound on the radius of heaviness kicks in, which again makes this event very unlikely.

Organization. We begin in [Section 3](#) with an overview of the numerous techniques we will require in our proof.

[Section 4](#) is dedicated to proving the mixing time bound for the diluted spin glass. In [Section 4.1](#), we describe a graph decomposition, and state several results that will play a part in the proof. In [Section 4.2](#), before proving any of these results, we begin by putting them together to prove an MLSI. In [Section 4.3](#), we bound the covariances of Ising models supported on near-forests using the trace method. In [Section 4.4](#), we then use this bound to prove a covariance bound that is used to show that entropy is conserved along the localization path. In [Section 4.5](#), we prove a modified log-Sobolev inequality for Ising models supported on near-forests, which we are left with at the end of the localization scheme.

Many of the results and ideas from the previous section make a return in [Section 5](#), where we prove an MLSI for Ising models whose interaction matrix is a scaling of a centered adjacency matrix. We begin with a high-level overview of the proof strategy in the centered setting. In [Section 5.3](#), we introduce some additional tools related to the spectral independence framework that are required for the proof, and then instantiate these tools to prove an MLSI for the Ising model obtained at the end of the localization scheme. In [Section 5.4](#), we reuse many of the tools introduced in the previous subsection to show that entropy is conserved along the path of the localization scheme.

Finally, in [Section 6](#), we prove that the decomposition of stochastic block models has the properties we claim.

3 Preliminaries

3.1 Notation

- Given a Markov chain P on Ω , and $x \in \Omega$, P_x is the distribution over Ω obtained by taking one step of the Markov chain from x .
- Given $\sigma \in \{\pm 1\}^n$, $\sigma^{\oplus i} \in \{\pm 1\}^n$ is defined as

$$\left(\sigma^{\oplus i}\right)_j = \begin{cases} \sigma_j, & j \neq i, \\ -\sigma_i, & j = i. \end{cases}$$

- Given a matrix M , we denote its operator norm by $\|M\| = \|M\|_{\text{op}}$.
- Given a distribution μ over Ω and a function $f : \Omega \rightarrow \mathbb{R}_{>0}$,

$$\text{Ent}[f] = \text{Ent}_{\mu}[f] := \mathbf{E}_{\mu}[f \log f] - \mathbf{E}_{\mu}[f] \log \mathbf{E}_{\mu}[f].$$

- For a random variable X over \mathbb{R}^k , we use $\mathbf{Cov}[X]$ to denote the covariance $\mathbf{E}[XX^{\top}] - \mathbf{E}[X]\mathbf{E}[X]^{\top}$. For a probability distribution μ over \mathbb{R}^k , we use $\mathbf{Cov}(\mu)$ to denote the covariance of the associated random variable with law μ .
- Given a connected graph $G = (V, E)$, its *tree-excess* is $|E| - |V| + 1$.
- Given a graph H , we use $V(H)$ and $E(H)$ to denote its vertex and edge sets respectively. We use $V'(H)$ to refer to the subset of vertices of $V(H)$ adjacent to at least one other distinct vertex, the set of non-isolated vertices.
- For a matrix $M \in \mathbb{R}^{n \times n}$ and a set of vertices $S \subseteq [n]$, by M_S we mean the $n \times n$ matrix where every row and column outside S is zeroed out, i.e., the principal submatrix of M restricted to S but preserving the dimension. If H is a graph on $[n]$, then by $M_H \in \mathbb{R}^{n \times n}$ we mean $M_{V'(H)}$, i.e., the restriction of M to the non-isolated vertices of H .
- We often abuse notation and identify probability measures μ with their density functions.
- Given a matrix C that is possibly not full-rank, and $z \sim \mathcal{N}(0, I)$, we mean by $C^{-1/2}z$ the distribution supported on $\text{Ker}(C)^{\perp}$ with density at z proportional to $\exp(-\frac{1}{2} \cdot z^{\top} C z)$.

3.2 Markov chains

In this section, we have a distribution μ over Ω , and shall look at how one can analyze Markov chains with stationary distribution μ . Throughout, we work solely with time-reversible Markov chains. First, we describe a natural Markov chain to sample from distributions over $\{\pm 1\}^n$.

Definition 3.1 (Glauber dynamics). Let $\Omega = \{\pm 1\}^n$, and suppose that μ is supported on all of Ω . The corresponding *Glauber dynamics* Markov chain $P = P_{\mu}$ is described by the following transition law from any state σ :

- Choose index i uniformly at random from $[n]$.
- Go to $\sigma^{\oplus i}$ with probability $\frac{\mu(\sigma^{\oplus i})}{\mu(\sigma) + \mu(\sigma^{\oplus i})}$, and stay at σ otherwise.

Fact 3.2. *The Markov chain P_{μ} is ergodic and time-reversible (with respect to μ), so has μ as its stationary distribution.*

Our goal is to bound the “mixing time” of Glauber dynamics on certain distributions.

Definition 3.3. Given a Markov chain P with stationary distribution μ , and a distribution ν that is absolutely continuous with respect to μ , the associated mixing time initialized at ν is defined by

$$t_{\text{mix}}(P, \varepsilon; \nu) = \min \{t > 0 : d_{\text{TV}}(P^t \nu, \mu) < \varepsilon\}.$$

We define the *mixing time* to be

$$t_{\text{mix}}(P, \varepsilon) = \max_{x \in \Omega} t_{\text{mix}}(P, \varepsilon; \delta_x).$$

It is well-known that the mixing time of Glauber dynamics (and Markov chains more generally) can be controlled by log-Sobolev inequalities.

Definition 3.4 (Dirichlet Form). Given a distribution μ over $\{\pm 1\}^n$, a time-reversible Markov chain P with stationary distribution μ , and functions $f, g : \{\pm 1\}^n \rightarrow \mathbb{R}$, the associated *Dirichlet form* is defined as

$$\mathcal{E}(f, g) = \mathcal{E}_P(f, g) := \mathbf{E}_{\substack{x \sim \mu \\ y \sim P_x}} [(f(x) - f(y))(g(x) - g(y))].$$

Definition 3.5 (Modified log-Sobolev Inequality). Given a time-reversible Markov chain P on $\{\pm 1\}^n$ with stationary distribution μ , P is said to satisfy a *modified log-Sobolev inequality* (MLSI) with constant C (possibly depending on n) if for any function $f : \{\pm 1\}^n \rightarrow \mathbb{R}_{>0}$,

$$\mathcal{E}(f, \log f) \geq C \text{Ent}[f].$$

In particular, C_{MLSI} is the best (largest) such constant C .

Fact 3.6 ([BT06, Corollary 2.8]). Given a time-reversible Markov chain P with stationary distribution μ , if P satisfies a modified log-Sobolev inequality with constant C_{MLSI} , then

$$t_{\text{mix}}(P, \varepsilon) \leq \frac{1}{C_{\text{MLSI}}} \left(\log \log \frac{1}{\mu_*} + \log \frac{1}{\varepsilon} \right),$$

where $\mu_* = \min_{x: \mu(x) > 0} \mu(x)$.

Fact 3.7 ([Goe04, Lemma 2.5]). Let μ be a distribution on $\{\pm 1\}^n$, and $(C_i)_{i=1}^k$ a partition of $[n]$ such that there exist distributions $\mu^{(i)}$ on $\{\pm 1\}^{C_i}$ with $\mu = \otimes_{i=1}^k \mu^{(i)}$. For the Glauber dynamics Markov chain, if each $\mu^{(i)}$ satisfies an MLSI with constant $\rho^{(i)}$, then μ satisfies an MLSI with constant $\rho = \frac{1}{n} \min_i \rho^{(i)} |C_i|$.

We will also require the Holley-Stroock perturbation principle, which shows that MLSIs and variances do not change too much if we multiply the densities by a constant.⁷

Lemma 3.8 ([HS86, p. 1186], see also [Led01, Lemma 1.2]). Let μ, ν be two probability measures over $\{\pm 1\}^n$, and suppose that for some $c > 1$ (possibly depending on n), $\frac{1}{c} \leq \frac{\nu(\sigma)}{\mu(\sigma)} \leq c$ uniformly over all $\sigma \in \{\pm 1\}^n$.

⁷ Their result works with log-Sobolev inequalities instead of modified log-Sobolev inequalities. Nevertheless, the proof is easily adapted.

1. Suppose Glauber dynamics for μ has MLSI constant at least C_{MLSI} . Then, Glauber dynamics for ν has MLSI constant at least C_{MLSI}/c^2 .

2. For any function f ,

$$\frac{1}{c} \cdot \mathbf{Var}_\mu[f] \leq \mathbf{Var}_\nu[f] \leq c \cdot \mathbf{Var}_\mu[f].$$

Also note that if $\nu \propto e^{-f}$ and $\mu \propto e^{-g}$, and $\|f - g\|_\infty \leq \tilde{c}$, then $e^{-2\tilde{c}} \leq \frac{\nu(\sigma)}{\mu(\sigma)} \leq e^{2\tilde{c}}$.

We make frequent use of the following simple observation about covariance matrices.

Observation 3.9. For any distribution μ on \mathbb{R}^n with finite second moment, and vector $v \in \mathbb{R}^n$,

$$v^\top \mathbf{Cov}(\mu)v = \mathbf{Var}[\langle v, x \rangle].$$

In particular,

$$\sup_{\substack{v \in \mathbb{S}^{n-1} \\ \text{supp}(v) \subseteq S}} \mathbf{Var}_{x \sim \mu}[\langle v, x \rangle] = \|\mathbf{Cov}(\mu)_S\|.$$

3.3 Ising models

In this section, we have a symmetric *interaction matrix* $J \in \mathbb{R}^{n \times n}$ and an *external field* $h \in \mathbb{R}^n$.

Definition 3.10 (Ising model). The *Ising model* corresponding to J and h is the probability distribution $\mu_{J,h}$ on $\{\pm 1\}^n$, where

$$\mu_{J,h}(\sigma) \propto \exp\left(\frac{1}{2}\sigma^\top J\sigma + \langle h, \sigma \rangle\right).$$

Its *partition function* is

$$Z_{J,h} := \sum_{\sigma \in \{\pm 1\}^n} \exp\left(\frac{1}{2}\sigma^\top J\sigma + \langle h, \sigma \rangle\right).$$

Observation 3.11. A simple observation we frequently make use of is that for any diagonal matrix D , $\mu_{J,h} = \mu_{J+D,h}$.

We rely on a well-known trick that lets us decompose an Ising model into a mixture of simpler Ising models.

Theorem 3.12 (Hubbard–Stratonovich transform). For any positive semidefinite matrix C in $\mathbb{R}^{n \times n}$:

$$\mu_{J,h} = \mathbf{E}_{z \sim \nu} \mu_{J-C, Cz+h}$$

where ν is a distribution over \mathbb{R}^n with density

$$\nu(z) \propto \exp\left(-\frac{1}{2}z^\top Cz + \log Z_{J-C, Cz+h}\right).$$

Further, $\mathbf{Cov}(\mu_{J,h}) \preceq \mathbf{Cov}(\nu)$.

We will only use the HS transform with C being a positive definite matrix.

Proof. Let $\sigma \sim \mu_{J,h}$ and $\mathbf{y} \sim \mathcal{N}(0, I)$ be independent, and define $\mathbf{z} := \sigma + C^{-1/2}\mathbf{y}$. The joint probability density function p of (z, σ) satisfies:

$$\begin{aligned} p(z, \sigma) &\propto \exp\left(-\frac{1}{2}(z - \sigma)^\top C(z - \sigma) + \frac{1}{2}\sigma^\top J\sigma + \langle h, \sigma \rangle\right) \\ &= \exp\left(-\frac{1}{2}z^\top Cz + \langle Cz, \sigma \rangle + \frac{1}{2}\sigma^\top (J - C)\sigma + \langle h, \sigma \rangle\right) \end{aligned}$$

This tells us:

$$\begin{aligned} \mu_{J,h}(\sigma) &\propto \int_{\mathbb{R}^n} p(z, \sigma) dz \\ &\propto \int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}z^\top Cz + \langle Cz, \sigma \rangle + \frac{1}{2}\sigma^\top (J - C)\sigma + \langle h, \sigma \rangle\right) dz \\ &= \int_{\mathbb{R}^n} \mu_{J-C, Cz+h}(\sigma) \cdot \exp\left(-\frac{1}{2}z^\top Cz + \log Z_{J-C, Cz+h}\right) dz, \end{aligned}$$

which completes the proof of the first part. The PSD inequality $\mathbf{Cov}(\mu_{J,h}) \preceq \mathbf{Cov}(v)$ follows from writing $\mathbf{Cov}(v)$ as $\mathbf{Cov}(\mu_{J,h}) + C^{-1}$. \blacksquare

The Brascamp–Lieb inequality is relevant to bounding the covariance matrix of distributions arising from the Hubbard–Stratonovich transform.

Theorem 3.13 (Brascamp–Lieb inequality). *For any distribution with density $v \propto \exp(-V)$ for some $V : \mathbb{R}^n \rightarrow \mathbb{R}$, if there is a matrix $\Gamma \succ 0$ such that $\nabla^2 V(z) \succeq \Gamma$ uniformly, then $\mathbf{Cov}(v) \preceq \Gamma^{-1}$.*

Using the Hubbard–Stratonovich transform with the Brascamp–Lieb inequality allows us to bound covariances using the following lemma.

Lemma 3.14. *Let M_1 and M_2 be interaction matrices in $\mathbb{R}^{n \times n}$, with M_1 positive definite, and let h be an external field in \mathbb{R}^n . Suppose for some positive definite matrix Γ , the following holds for all $z \in \mathbb{R}^n$:*

$$M_1 - M_1 \cdot \mathbf{Cov}(\mu_{M_2, M_1 z + h}) \cdot M_1 \succeq \Gamma.$$

Then $\mathbf{Cov}(\mu_{M_1 + M_2, h}) \preceq \Gamma^{-1}$.

Proof. First, observe that by applying [Theorem 3.12](#) with $C = M_1$, we obtain $\mathbf{Cov}(\mu_{M_1 + M_2, h}) \preceq \mathbf{Cov}(v)$ where $v(z) \propto \exp(-\frac{1}{2}z^\top M_1 z + \log Z_{M_2, M_1 z + h})$. It then suffices to bound $\mathbf{Cov}(v)$. Towards doing so, we show that v is log-concave, and then use the Brascamp–Lieb inequality ([Theorem 3.13](#)).

We can verify that:

$$-\nabla^2 \log v(z) = M_1 - \nabla^2 \log Z_{M_2, M_1 z + h} = M_1 - M_1 \cdot \mathbf{Cov}(\mu_{M_2, M_1 z + h}) \cdot M_1.$$

By assumption, for any $z \in \mathbb{R}^n$, $-\nabla^2 \log v(z) \succeq \Gamma \succ 0$, which establishes that v is log-concave. Consequently, by [Theorem 3.13](#) we get $\mathbf{Cov}(v) \preceq \Gamma^{-1}$. \blacksquare

To bound the above quantity, the following lemma will be useful.

Fact 3.15. *Let L, M be symmetric matrices such that $\|M\| \leq \alpha$ and $0 \prec \eta_1 \preceq L \preceq \frac{1-\eta_2}{\alpha}$. Then,*

$$L - LML \succeq \eta_1 \eta_2.$$

3.4 Localization schemes

We will make use of *stochastic localization* to relate our Markov chain to simpler Markov chains [Eld13, CE22]. We use a restricted form of stochastic localization, that we state below.

Definition 3.16 (Stochastic localization for Ising models). A *stochastic localization scheme* is a stochastic process $(\mu_t)_{0 \leq t \leq 1}$ on the space of Ising models that is parameterized by a (possibly time-varying) control matrix $(C_t)_{0 \leq t \leq 1}$ and an initialization $\mu_{J,h}$, described by the following stochastic differential equation.

$$d\left(\frac{\mu_t}{\mu_{J,h}}\right) = \frac{\mu_t}{\mu_{J,h}} \langle x - \mathbf{E}_{\mu_t} x, C_t \cdot dB_t \rangle \quad (3)$$

where $(B_t)_{0 \leq t \leq 1}$ is the standard Brownian motion in n dimensions.

While (3) may appear rather mysterious at first, the measures μ_t take on a natural form.

Fact 3.17 ([CE22, Facts 13 and 14]). *The distribution μ_t is the Ising model μ_{J_t, h_t} where $J_t = J - \int_0^t C_s^2 ds$ and $h_t = \int_0^t C_s \cdot dB_s + C_s^2 \mathbf{E}_{\mu_s}[x] \cdot ds$. Further, μ_t is a martingale, in that for any set $A \subseteq \{\pm 1\}^n$, $\mathbf{E}[\mu_t(A) | \mu_s] = \mu_s(A)$ for $0 \leq s \leq t \leq 1$.*

The following is a mild generalization of [CE22, Proposition 39 and Lemma 40] that tells us entropy conservation for stochastic localization; this stronger version is implicitly proved in [CE22, Lemma 40].

Lemma 3.18. *Consider a stochastic localization process $(\mu_t)_{t \geq 0}$ with respect to (possibly time-varying) control matrix $(C_t)_{t \geq 0}$, started at an Ising model $\mu_0 = \mu_{J,h}$. Suppose that for every $h' \in \mathbb{R}^n$ and $t \in [0, 1]$,*

$$\|C_t \cdot \mathbf{Cov}(\mu_{J_t, h'}) \cdot C_t\| \leq \alpha_t.$$

Then, for any positive-valued φ ,

$$\mathbf{E} [\text{Ent}_{\mu_1}[\varphi]] \geq \exp\left(-\int_0^1 \alpha_t dt\right) \cdot \text{Ent}_{\mu_{J,h}}[\varphi].$$

The following result articulates how the measure decomposition provided by stochastic localization is instrumental in proving a modified log-Sobolev inequality for the Glauber dynamics Markov chain.

Theorem 3.19 ([CE22, Theorem 47]). *Let J be an interaction matrix and h be an external field. Let $(\mu_t)_{0 \leq t \leq 1}$ be a stochastic localization scheme initialized at $\mu_{J,h}$. Suppose that*

1. *We have the entropy conservation inequality*

$$\frac{\mathbf{E}[\text{Ent}_{\mu_1}[f]]}{\text{Ent}_{\mu_{J,h}}[f]} \geq \varepsilon$$

for all $f : \Omega \rightarrow \mathbb{R}_+$.

2. *Almost surely,*

$$C_{\text{MLSI}}(P_{\mu_1}) \geq \delta.$$

Then, $\mu_{J,h}$ has an MLSI coefficient at least $\varepsilon\delta$.

Using the above two results easily yields the following.

Theorem 3.20. Let M_1, M_2 be matrices in $\mathbb{R}^{n \times n}$ with M_1 positive semidefinite, and suppose that for every external field $h' \in \mathbb{R}^n$ and $t \in [0, 1]$,

1. $\left\| M_1^{1/2} \cdot \mathbf{Cov}(\mu_{(1-t)M_1 + tM_2, h'}) \cdot M_1^{1/2} \right\| \leq q$, and
2. Glauber dynamics for $\mu_{M_2, h'}$ satisfies $C_{\text{MLSI}}(\mu_{M_2, h'}) \geq \delta$,

then, for any external field $h \in \mathbb{R}^n$,

$$C_{\text{MLSI}}(\mu_{M_1 + M_2, h}) \geq \delta \cdot e^{-q}.$$

Proof. We will use [Theorem 3.19](#) by choosing the stochastic localization to be that with control matrix $M_1^{1/2}$. The desired statement follows once we verify that the following are true:

1. $\frac{\mathbf{E} [\text{Ent}_{\mu_1}[f]]}{\text{Ent}_{\mu_{J,h}}[f]} \geq e^{-q}$ for every $f : \{\pm 1\}^n \rightarrow \mathbb{R}_+$, and
2. $C_{\text{MLSI}}(\mu_1) \geq \delta$ almost surely.

[Item 1](#) follows from [Lemma 3.18](#), and [Item 2](#) is true by assumption, which completes the proof. ■

3.5 Graphs

For the rest of this section, let G be a graph on n vertices and m edges, and let c be a weight function on this graph. We will use A_c to denote the adjacency matrix of G , weighted according to c , and D_c as the weighted degree function where $D_c[i, i] = \sum_{j \neq i} A_c[i, j]$.

Definition 3.21 (Nonbacktracking matrix). The *nonbacktracking matrix* B_c is a $2m \times 2m$ matrix indexed by directed edges in G where:

$$B_c[uv, xy] = \begin{cases} c(xy) & \text{if } v = x \text{ and } u \neq y \\ 0 & \text{otherwise.} \end{cases}$$

We will use B_G to refer to the *unweighted* nonbacktracking matrix, i.e., when $c(xy)$ is equal to 1 for every edge in G .

The nonbacktracking matrix is related to a matrix known as the Bethe Hessian, and we use this relationship to obtain information about the spectrum of the adjacency matrices of bounded degree subgraphs of a random graph arising from the stochastic block model.

Definition 3.22 (Bethe Hessian). The *Bethe Hessian* is defined as: $\text{BH}_c := D_{c\hat{c}} - A_{\hat{c}} + I$ where $\hat{c} = \frac{c}{1-c^2}$. When all the weights are equal to t , the Bethe Hessian can be written as:

$$\text{BH}_G(t) := \frac{(D_G - I)t^2 - At + I}{1 - t^2}.$$

The following is a classic statement that relates the eigenvalues of the Bethe Hessian and those of the nonbacktracking matrix; see, e.g., [FM17, Proof of Theorem 5.1]).

Lemma 3.23. *For any $\alpha \in (0, 1)$, the spectral radius $\rho(B_G) \leq \frac{1}{\alpha}$ if and only if $\text{BH}_G(t) \succ 0$ for all $t \in (-\alpha, \alpha)$.*

A simple observation is that when G is a tree, its nonbacktracking matrix is nilpotent, and hence its spectral radius is equal to 0, which gives us the following consequence.

Corollary 3.24. *When G is a tree, for all $t \in (-1, 1)$, we have $\text{BH}_G(t) \succ 0$.*

We define $A_G^{(s)}$ as the $n \times n$ matrix where the i, j entry contains the number of length- s nonbacktracking walks between i and j . The following is a classic fact; see, e.g., [ST96, Equation (2.1)].

Fact 3.25. $\text{BH}_G(t)^{-1} = \sum_{s \geq 0} A_G^{(s)} t^s$.

We will need to understand the spectrum of the nonbacktracking matrix of a random graph drawn from the stochastic block model, as well as random graphs with randomly signed edge weights. For the sequel, let $G \sim \text{SBM}(n, d, \lambda)$ with hidden community vector $\sigma \in \{\pm 1\}^n$, and assume $d > 1$. Consider the adjacency matrix centered after conditioning on σ , i.e., the matrix $\overline{A}_G := A_G - \mathbf{E}[A_G | \sigma]$. Let c be the weight function on the edges of the complete graph prescribed by \overline{A}_G .

The following bounds on the nonbacktracking matrix of a random graph drawn from the stochastic block model are effectively due to [BLM15], but the exact form we will need is the version in [LMR22].

Lemma 3.26 (Consequence of [LMR22, Theorem 7.4]). *For $\varepsilon > 0$:*

$$|\lambda|_{\max}(B_c) \leq (1 + \varepsilon) \cdot \sqrt{d}$$

with probability $1 - o_n(1)$.

By an identical argument to the proof of [FM17, Theorem 5.1], the following is true.

Lemma 3.27. *For any $t \in \left(-\frac{1}{\sqrt{d}}, \frac{1}{\sqrt{d}}\right)$, $\text{BH}_{c,t} \succeq 0$ with probability $1 - o_n(1)$.*

Recalling that the centered adjacency matrix is given by $\overline{A}_G = A_G - \mathbf{E}A_G | \sigma$, and using the fact that the entries of $\mathbf{E}A_G | \sigma$ are of magnitude $O\left(\frac{d}{n}\right)$, and the fact that all degrees in G are at most $O(\log n)$ with high probability, we get the following relationship.

Corollary 3.28. *For any $t \in \left(-\frac{1}{\sqrt{d}}, \frac{1}{\sqrt{d}}\right)$, we get that with probability $1 - o_n(1)$:*

$$(D_G - I)t^2 - A_G t + \mathbf{E}[A_G | \sigma] \cdot t(1 - t^2) + I(1 + o(1)) \succeq 0.$$

Further, by using the fact that $\|\mathbf{E}[A_G | \sigma]\| \leq d \cdot (1 + o(1))$, we get:

$$\widetilde{\text{BH}}_G(t) := (D_G - I)t^2 - \overline{A}_G t + I \cdot (1 + d|t|^3 + o(1)) \succeq 0.$$

We are finally ready to prove the main result about the spectrum of the bulk in our decomposition.

Lemma 3.29. *Let $S \subseteq [n]$ such that the degree of every vertex in S within \mathbf{G} is at most $(1 + \varepsilon)d$, and suppose that $\sqrt{d(1 + \varepsilon)} + \frac{1}{1 + \varepsilon} \leq \sqrt{d}(1 + \varepsilon)$. Then*

$$\|(\overline{A}_{\mathbf{G}})_S\| \leq 2\sqrt{d}(1 + \varepsilon)$$

with probability $1 - o_n(1)$. Consequently, the spectral diameter of $(\overline{A}_{\mathbf{G}})_S$ is at most $4\sqrt{d}(1 + \varepsilon)$.

Proof. We shall use [Corollary 3.28](#). Set $t = \frac{1}{(1 + \varepsilon)\sqrt{d}}$. Observing that any principal submatrix of the PSD matrix $\widetilde{\mathbf{B}}\mathbf{H}_{\mathbf{G}}(t) \succeq 0$ is also PSD, and $\|(D_{\mathbf{G}})_S\| \leq d(1 + \varepsilon)$, we get that

$$\begin{aligned} \lambda_{\max}((\overline{A}_{\mathbf{G}})_S) &\leq \frac{1 + dt^3 - t^2 + d(1 + \varepsilon)t^2}{t} \\ &\leq \sqrt{d}(1 + \varepsilon) + \frac{1}{1 + \varepsilon} + \sqrt{d(1 + \varepsilon)} \leq 2\sqrt{d}(1 + \varepsilon). \end{aligned}$$

Doing the same for the matrix $\widetilde{\mathbf{B}}\mathbf{H}_{\mathbf{G}}(-t) \succeq 0$,

$$\begin{aligned} \lambda_{\min}((\overline{A}_{\mathbf{G}})_S) &\geq \frac{1 + dt^3 - t^2 + d(1 + \varepsilon)t^2}{-t} \\ &\geq -\sqrt{d}(1 + \varepsilon) - \frac{1}{1 + \varepsilon} - \sqrt{d(1 + \varepsilon)} \geq -2\sqrt{d}(1 + \varepsilon). \quad \blacksquare \end{aligned}$$

Similarly, there are well-established bounds on the spectral radius of the nonbacktracking matrix of randomly signed Erdős–Rényi graphs.

Lemma 3.30 (Consequence of [\[SM22, Theorem 2\]](#)). *Let $\mathbf{G} \sim \text{SBM}(n, d, \lambda)$ and let \mathbf{c} be a random assignment of ± 1 -valued signs to the edges of \mathbf{G} . Then, with probability $1 - o_n(1)$,*

$$|\lambda|_{\max}(B_{\mathbf{c}}) \leq (1 + o_n(1)) \cdot \sqrt{d}.$$

An identical argument to the proof of [Lemma 3.29](#) can be used to establish the following.

Lemma 3.31. *Let $\mathbf{G} \sim \text{SBM}(n, d, \lambda)$ and let \mathbf{c} be a random assignment of ± 1 -valued signs to the edges of \mathbf{G} . With probability $1 - o_n(1)$, the following holds. Let $S \subseteq [n]$ such that the degree of every vertex in S within \mathbf{G} is at most $(1 + \varepsilon)d$. Then,*

$$\|A_{\mathbf{c}}\| \leq 2\sqrt{d}(1 + \varepsilon).$$

Remark 3.32. Notice that one difference between [Lemma 3.29](#) and [Lemma 3.31](#) is that the latter does not require a constraint on d and ε . This owes to the fact that in the former case there are nonzero entries both at the scale of $\Theta(1)$ and $\Theta(1/n)$ whereas all nonzero entries in the latter case are in $\{\pm 1\}$.

4 Glauber dynamics on near-forest decompositions

This section is dedicated to proving the following result.

Theorem 4.1. *Fix $d > 1$. Let A be the adjacency matrix of a randomly signed graph $G \sim \text{SBM}(n, d, \lambda)$. There exists a universal constant $\beta > 0$ such that with high probability over the draw of G , for any external field h , we have $C_{\text{MLSI}}(\mu_{\beta A/\sqrt{d}, h}) \geq n^{-1-o(1)}$.*

Given a sparse random graph G , our goal is to show that for any external field h , Glauber dynamics for the randomly signed Ising model on G mixes rapidly in the appropriate temperature regime.

In [Section 4.1](#), we describe the graph decomposition, and state several results that will play a part in the proof. In [Section 4.2](#), before proving any of these results, we begin by putting them together to prove an MLSI — the localization scheme we use will be described therein. In [Section 4.3](#), we bound the covariances of Ising models supported on near-forests using the trace method — these bounds are used in [Section 4.4](#), where we establish a covariance bound used in showing that entropy is conserved along the localization path. In [Section 4.5](#), we prove a modified log-Sobolev inequality for Ising models supported on near-forests, which we are left with at the end of the localization scheme.

4.1 On decomposing random graphs

In this section, we describe an explicit class of decompositions of an Ising Hamiltonian into the sum of two Hamiltonians which are supported on two disjoint sets of edges, which we term the *bulk* and the *near-forest* components. We then use a localization scheme to anneal the original Ising model into an Ising model with interactions supported only on the near-forest component. We will verify the two conditions needed to apply [Theorem 3.20](#) to imply rapid mixing of Glauber dynamics on sparse random graphs.

Definition 4.2 (Near-forest). We say a graph H is a *near-forest* if all its connected components have at most a single cycle. Equivalently, each component has tree-excess at most 1.

A crucial ingredient in our proof is that sparse stochastic block model graphs can be decomposed into the union of two graphs, a *bulk* and a *near-forest*, where the bulk has bounded degree and the near-forest satisfies some pseudorandom properties. The following definitions capture the pseudorandom property the near-forest satisfies and the precise form of the decomposition.

Definition 4.3 (Pseudorandom graph). We say a graph H is (Δ, D) -*pseudorandom* with respect to $S \subseteq V(H)$ if it satisfies the following:

1. For any $r \geq 1$, and any vertex $u \in V(H)$, the size of the r -ball around u is at most $D^{r-1} \cdot \Delta$. Note in particular that this implies that the maximum degree is at most Δ .
2. For any $r \geq 0$ and $u \in V(H)$, $|\{v \in S : \text{dist}(u, v) \leq r\}| \leq D^r$.

While the above statement bounds the sizes of balls, most of our proofs will only need bounds on the sizes of neighborhoods. We are now ready to formally state our graph decomposition.

Definition 4.4 (Graph decomposition). For a graph G , let B, H be subgraphs of G with possibly overlapping sets of vertices, which we refer to as the *bulk* and *near-forest* components. We say that (B, H) is a (Δ, D) -decomposition of G if the following properties hold.

1. $E(G) = E(B) \sqcup E(H)$, i.e. the bulk and near-forest components partition the edges of G .
2. Let $\partial H = V'(B) \cap V'(H)$ denote the vertex boundary of H , the vertices in H incident to an edge in $E(B)$. Then H is a (Δ, D) -pseudorandom near-forest with respect to ∂H .

If such a (B, H) exists, we say that G admits a (Δ, D) -decomposition.

Many of our results hold generically assuming that G admits a (Δ, D) -decomposition (or is itself a (Δ, D) -pseudorandom near-forest); we refer the readers back to [Figure 1](#) for an illustration of such a decomposition. In [Theorem 4.9](#), we show that sparse stochastic block model graphs admit such a decomposition.

As a crucial first step towards analyzing the class of Ising models supported on such graphs, we establish the following covariance bounds for *any* Ising model with bounded interactions supported on a pseudorandom near-forest; we defer the proof to [Section 4.3](#).

Lemma 4.5. *Let H be a (Δ, D) -pseudorandom near-forest with respect to $S \subseteq V(H)$. For any Ising model with interactions J supported on $E(H)$ with values in $\left[-\frac{\gamma}{\sqrt{D}}, \frac{\gamma}{\sqrt{D}}\right]$, along with an arbitrary external field h , we have*

$$\|\mathbf{Cov}(\mu_{J,h})_S\| \leq \frac{e^{2\gamma/\sqrt{D}}}{(1-\gamma)^2}.$$

Additionally,

$$\|\mathbf{Cov}(\mu_{J,h})\| \leq \frac{e^{2\gamma/\sqrt{D}}}{(1-\gamma)^2} \cdot \frac{\Delta}{D}.$$

As a consequence of [Lemma 4.5](#), we can establish conservation of entropy and a good MLSI constant for the annealed model. We prove these implications in [Sections 4.4](#) and [4.5](#), respectively.

Theorem 4.6. *Let (B, H) be a (Δ, D) -decomposition of G , and let $J_B, J_H \in \mathbb{R}^{n \times n}$ be interaction matrices whose off-diagonal entries are supported on $E(B)$ and $E(H)$, respectively. Suppose that there are constants $\eta, K \in (0, 1)$ such that $\eta I_B \preceq J_B \preceq K I_B$, and that all interactions in J_H are in $\left[-\frac{\gamma}{\sqrt{D}}, \frac{\gamma}{\sqrt{D}}\right]$ for some $\gamma \in [0, 1)$. Finally, suppose that*

$$K \cdot \frac{e^{2\gamma/\sqrt{D}}}{(1-\gamma)^2} \leq 1 - \delta$$

for a positive constant $\delta > 0$. There exists a positive constant C not depending on Δ such that for any $0 \leq t \leq 1$ and external field $h \in \mathbb{R}^n$,

$$\mathbf{Cov}(\mu_{(1-t)J_B + tJ_H, h}) \preceq C(I_B + \Delta I_{H \setminus \partial H}).$$

Remark 4.7. The assumptions on the spectrum of J_B may look artificial, but they are in fact without loss of generality. Indeed, by [Observation 3.11](#), one can shift J_B by a constant multiple of I_B to ensure it is PSD lower bounded by η when restricted to the bulk.

Theorem 4.8. Let H be a (Δ, D) -pseudorandom near-forest, and $J \in \mathbb{R}^{n \times n}$ an interaction matrix supported on H with all interactions in $\left[-\frac{\gamma}{\sqrt{D}}, \frac{\gamma}{\sqrt{D}}\right]$ for some $\gamma \in [0, 1)$. There is a constant c (depending on γ) such that for any external field $h \in \mathbb{R}^n$,

$$C_{\text{MLSI}}(\mu_{J,h}) \geq \frac{1}{ne^{c\Delta}}.$$

To apply [Theorems 4.6](#) and [4.8](#) for concrete graphs to prove rapid mixing results, we need to establish that these graphs admit a (Δ, D) -decomposition.

For sparse random graphs, we explicitly construct the near-forest decomposition as outlined in [Section 2](#) and obtain the following explicit bounds on the parameters Δ and D ; the analysis is given in [Section 6](#).

Theorem 4.9. Let $\mathbf{G} \sim \text{SBM}(n, d, \lambda)$ for any $d \geq 1$. There exist universal constants C_1 and C_2 such that for any $\varepsilon \geq C_1 \cdot \left(\frac{C_2 + \log d}{d}\right)^{1/3}$, with high probability over \mathbf{G} , we can partition the edge set $E(\mathbf{G})$ as $E(\mathbf{G}_1) \sqcup E(\mathbf{G}_2)$ for graphs \mathbf{G}_1 and \mathbf{G}_2 on $[n]$ where:

1. The maximum degree of \mathbf{G}_1 is at most $(1 + \varepsilon)d$.
2. \mathbf{G}_2 is a (Δ, D) -pseudorandom near-forest with respect to $S := V'(\mathbf{G}_1) \cap V'(\mathbf{G}_2)$ for $\Delta = o(\log n)$ and $D = (1 + \varepsilon)d$.

Remark 4.10. We will henceforth refer to \mathbf{G}_1 as the *bulk* part in the decomposition, and to \mathbf{G}_2 as the *near-forest* part of the decomposition.

Remark 4.11. The bound on the maximum degree of \mathbf{G}_1 is important for obtaining a bound on the spectral norm of the interaction matrix restricted to the edges in \mathbf{G}_1 . The bound on Δ of $o(\log n)$ is crucial to obtain a bound of $n^{1+o(1)}$ on the mixing time via [Theorem 4.8](#).

4.2 Proving an MLSI for diluted spin glasses

In this section, we apply [Theorems 4.6](#), [4.8](#) and [4.9](#) to obtain a lower bound on the MLSI constant for Glauber dynamics on randomly signed sparse random graphs. We restate the theorem here for convenience.

Theorem 4.1. Fix $d > 1$. Let A be the adjacency matrix of a randomly signed graph $\mathbf{G} \sim \text{SBM}(n, d, \lambda)$. There exists a universal constant $\beta > 0$ such that with high probability over the draw of \mathbf{G} , for any external field h , we have $C_{\text{MLSI}}(\mu_{\beta A/\sqrt{d}, h}) \geq n^{-1-o(1)}$.

Proof. Consider the bulk \mathbf{G}_1 and near-forest \mathbf{G}_2 guaranteed by [Theorem 4.9](#), and let $J_B = \frac{\beta}{\sqrt{d}}A_{\mathbf{G}_1}$ and $J_H = \frac{\beta}{\sqrt{d}}A_{\mathbf{G}_2}$. We will verify the conditions of [Theorem 4.6](#) for some $\beta = \Omega(1)$. We want to verify that

$$K \cdot \frac{e^{2\gamma/\sqrt{D}}}{(1-\gamma)^2} \leq 1 - \delta \tag{4}$$

Before we dive into the calculations, let us first observe that from [Theorem 4.9](#), we have that \mathbf{G}_2 is a $(o(\log n), d(1 + \varepsilon))$ -pseudorandom near-forest. Thus, the interaction strengths on J_H are in $\left[-\frac{\beta}{\sqrt{d}}, \frac{\beta}{\sqrt{d}}\right] = \left[-\frac{\beta\sqrt{1+\varepsilon}}{\sqrt{d(1+\varepsilon)}}, \frac{\beta\sqrt{1+\varepsilon}}{\sqrt{d(1+\varepsilon)}}\right]$. We will thus check (4) for $\gamma = \beta\sqrt{1 + \varepsilon}$. We have

$$\frac{e^{2\gamma/\sqrt{D}}}{(1 - \gamma)^2} = \frac{e^{2\beta/\sqrt{d}}}{(1 - \beta\sqrt{1 + \varepsilon})^2}.$$

Turning now to K , [Lemma 3.31](#) furnishes the bound

$$-2\beta(1 + \varepsilon)I_B \preceq J_B \preceq 2\beta(1 + \varepsilon)I_B.$$

For some “small” $\eta > 0$ that we fix later, we can define $\tilde{J}_B = J_B + (\eta + 2\beta(1 + \varepsilon))I_B$ without changing the Ising model to ensure

$$\eta I_B \preceq \tilde{J}_B \preceq (\eta + 4\beta(1 + \varepsilon))I_B.$$

Thus, we choose $K = \eta + 4\beta(1 + \varepsilon)$. Hence, the LHS of (4) is at most

$$K \cdot \frac{e^{2\beta/\sqrt{d}}}{(1 - \beta\sqrt{1 + \varepsilon})^2} \leq e^{2\beta/\sqrt{d}} \cdot \frac{\eta + 4\beta(1 + \varepsilon)}{(1 - \beta(1 + \varepsilon))^2}.$$

If we have

$$\beta < \min \left\{ 1, \frac{1 + 2e^{2/\sqrt{d}} - \sqrt{4e^{4/\sqrt{d}} + 4e^{2/\sqrt{d}}}}{1 + \varepsilon} \right\} = (3 - \sqrt{8})(1 - o_d(1)) \approx 0.17 \cdot (1 - o_d(1)),$$

sufficiently small constants η and δ may be chosen so the above is less than $1 - \delta$.

Now, we can apply [Theorem 4.6](#) to show that

$$\mathbf{Cov}(\mu_{(1-t)J_B+J_H,h}) \preceq C(I_B + o(\log n) \cdot I_{H \setminus \partial H}),$$

and in turn

$$\exp(-\|\tilde{J}_B^{1/2} \cdot \mathbf{Cov}(\mu_{(1-t)J_B+J_H,h}) \cdot \tilde{J}_B^{1/2}\|) \geq \Omega(1).$$

On the other hand, [Theorem 4.8](#) shows that for any h' we have

$$C_{\text{MLSI}}(\mu_{J_H,h'}) \geq \frac{1}{n} \cdot e^{-o(\log n)} = n^{-1-o(1)}.$$

Finally, applying [Theorem 3.20](#) with $M_1 = \tilde{J}_B$ and $M_2 = J_H$ allows us to conclude that $C_{\text{MLSI}}(\mu_{\beta A/\sqrt{d},h}) \geq n^{-1-o(1)}$. ■

Combined with [Fact 3.6](#), we immediately conclude [Theorem 1.1](#).

4.3 Covariance bounds on near-forests

In this section we establish [Lemma 4.5](#), which bounds the covariance of an Ising model supported on a pseudorandom near-forest.

Lemma 4.5. *Let H be a (Δ, D) -pseudorandom near-forest with respect to $S \subseteq V(H)$. For any Ising model with interactions J supported on $E(H)$ with values in $\left[-\frac{\gamma}{\sqrt{D}}, \frac{\gamma}{\sqrt{D}}\right]$, along with an arbitrary external field h , we have*

$$\|\mathbf{Cov}(\mu_{J,h})_S\| \leq \frac{e^{2\gamma/\sqrt{D}}}{(1-\gamma)^2}.$$

Additionally,

$$\|\mathbf{Cov}(\mu_{J,h})\| \leq \frac{e^{2\gamma/\sqrt{D}}}{(1-\gamma)^2} \cdot \frac{\Delta}{D}.$$

Our strategy to prove [Lemma 4.5](#) is:

1. Reduce to proving the covariance norm bound in the case when H is a tree, the interactions are nonnegative, and there is no external field.
2. Use an explicit formula for the covariance in tree Ising models with a zero external field, and bound the spectral norm of that matrix using the trace moment method.

The first ingredient in this reduction is the following simple observation.

Observation 4.12. For any interaction matrix J , external field h , and diagonal matrix of signs D , we have:

$$\|\mathbf{Cov}(\mu_{J,h})\| = \|\mathbf{Cov}(\mu_{DJD,Dh})\|.$$

In fact, $C_{\text{MLSI}}(\mu_{J,h}) = C_{\text{MLSI}}(\mu_{DJD,Dh})$.

Proof. Note that if $\sigma \sim \mu_{J,h}$, the vector $D\sigma$ is distributed as $\mu_{DJD,Dh}$. Thus, $\mathbf{Cov}(\mu_{DJD,Dh}) = D \cdot \mathbf{Cov}(\mu_{J,h}) \cdot D$. The first observation now follows since D is an orthogonal matrix. For the second part, we have that for any function $f : \{\pm 1\}^n \rightarrow \mathbb{R}_{>0}$, if we define $g : \{\pm 1\}^n \rightarrow \mathbb{R}_{>0}$ by $g(\sigma) = f(D\sigma)$, then the entropies and Dirichlet forms of f and g with respect to $\mu_{J,h}$ and $\mu_{DJD,Dh}$ (respectively) are equal. ■

We now use the above observation to pass from an Ising model on a forest with arbitrary interactions to one with nonnegative interactions on all but one edge.

Lemma 4.13. *For any connected graph H and spanning tree T of H , interaction matrix J supported on H , and an external field h , there is a diagonal matrix D such that DJD has nonnegative interactions on the edges of T .*

Proof. Arbitrarily root T at a vertex r . Construct D by choosing $D_{v,v} = \prod_{e' \in P} \text{sign}(J_{e'})$ where P is the unique path from r to v in T . For any edge $uv \in E(T)$, $(DJD)_{u,v} = J_{u,v} D_{u,u} D_{v,v} = J_{u,v} \text{sign}(J_{u,v}) \geq 0$. ■

We will need the following fact about ferromagnetic Ising models with external fields.

Lemma 4.14 ([DSS23, Corollary 1.3]). *For any ferromagnetic interaction matrix $J \in \mathbb{R}_{\geq 0}^{n \times n}$ and external field $h \in \mathbb{R}^n$, we have for any $u, v \in [n]$:*

$$\mathbf{Cov}(\mu_{J,h})_{u,v} \leq \mathbf{Cov}(\mu_{J,0})_{u,v}.$$

Fact 4.15. *For any ferromagnetic interaction matrix $J \in \mathbb{R}_{\geq 0}^{n \times n}$, external field $h \in \mathbb{R}^n$, and $u, v \in [n]$, $\mathbf{Cov}(\mu_{J,h})_{u,v} \geq 0$.*

Proof. Ferromagnetic Ising models with arbitrary external fields satisfy the hypothesis of the FKG inequality [Gri06, Theorem 2.16]. The statement then follows from the FKG inequality applied to the functions $\sigma_i - \mathbf{E}[\sigma_i]$ and $\sigma_j - \mathbf{E}[\sigma_j]$. \blacksquare

Corollary 4.16. *Let M be a matrix with all non-negative entries. For any ferromagnetic interaction matrix $J \in \mathbb{R}_{\geq 0}^{n \times n}$ and external field $h \in \mathbb{R}^n$, we have*

$$\|\mathbf{Cov}(\mu_{J,h}) \cdot M\| \leq \|\mathbf{Cov}(\mu_{J,0}) \cdot M\|.$$

In particular,

$$\|\mathbf{Cov}(\mu_{J,h})\| \leq \|\mathbf{Cov}(\mu_{J,0})\|.$$

Proof. This is immediate by Fact 4.15 and Lemma 4.14, and the Perron-Frobenius theorem. \blacksquare

Next, we prove a bound on the operator norm of matrices that encode when two vertices are a certain distance apart.

Lemma 4.17. *Let H be a (Δ, D) -pseudorandom tree with respect to $S \subseteq V(H)$. Let $\ell \geq 1$, and set $A^{(\ell)} = A_H^{(\ell)}$ to be the matrix with $(A^{(\ell)})_{uv} = 1$ if vertices u and v are distance exactly ℓ apart, and 0 otherwise. Also let $A = A^{(1)}$ be the adjacency matrix of H . Then,*

$$\begin{aligned} \|A^{(\ell)}\| &\leq D^{\ell/2} \cdot (\ell + 1) \cdot \left(\frac{\Delta}{D}\right) \\ \|A_S^{(\ell)}\| &\leq D^{\ell/2} \cdot (\ell + 1) \\ \|A^{(\ell)}A\| &\leq D^{\ell/2} \cdot (\ell + 1) \cdot \left(\frac{2\Delta}{\sqrt{D}}\right). \end{aligned}$$

Proof. We shall bound these operator norms using the trace moment method. Let R be S or $V(H)$, so we are interested in bounding $\|A_R^{(\ell)}\|$. Let us define $\mathcal{W}_{\ell,m,R}$ to be the set of all closed walks in H of length $\ell \cdot m$ that can be expressed as the concatenation of m nonbacktracking walks⁸ between a pair of vertices in R of length- ℓ each. Henceforth, we refer to each of these length- ℓ nonbacktracking walks as a *linkage*. For any even integer $m > 0$, we have:

$$\|A_R^{(\ell)}\|^m \leq \text{Tr}\left(\left(A_R^{(\ell)}\right)^m\right) = |\mathcal{W}_{\ell,m,R}|.$$

For each vertex $r \in R$, we count the number of walks that start and end at r . The t -th linkage, if started at a vertex v , is composed of s_t steps “upwards” away from v , followed by $\ell - s_t$ steps

⁸ In a tree, nonbacktracking walks are equivalent to shortest paths.

“downwards”. Since H is a tree, and the walks we are counting start and end at the same vertex, exactly half of the steps in any walk are upwards, and exactly half are downwards, that is,

$$\sum_{t=1}^m s_t = \sum_{t=1}^m \ell - s_t = \ell \cdot m/2.$$

Given s_t , the number of choices for the t -th linkage is at most $D^{\ell-s_t-1} \cdot \Delta_R$ where Δ_R is D when $R = S$ and Δ when $R = V(H)$.

$$|\mathcal{W}_{\ell,m,R}| \leq |R| \cdot \sum_{s_1, \dots, s_m} \prod_{t=1}^m D^{\ell-s_t} \cdot \frac{\Delta_R}{D} = |R| \cdot \sum_{s_1, \dots, s_m} D^{\ell \cdot m/2} \cdot \left(\frac{\Delta_R}{D}\right)^m \leq |R| \cdot D^{\ell \cdot m/2} \cdot \left(\frac{\Delta_R}{D}\right)^m \cdot (\ell+1)^m.$$

Thus, taking $m \rightarrow \infty$ so $|R|^{1/m} \rightarrow 1$, we get that

$$\|A_R^{(\ell)}\| \leq D^{\ell/2} \cdot \left(\frac{\Delta_R}{D}\right) \cdot (\ell+1).$$

The first two inequalities follow by plugging in $\Delta_S = D$ and $\Delta_{V(H)} = \Delta$.

The proof of the third inequality is very similar. The main difference is that each of the linkages is not a length- ℓ nonbacktracking walk, but instead a length- $(\ell+1)$ walk, of which the first ℓ steps are non-backtracking. That is, the t -th “linkage” is now composed of s_t steps upwards, followed by $\ell - s_t$ steps downwards, followed by a single step upwards or downwards. If the final step is downwards, this is just s_t steps upwards then $\ell - s_t + 1$ steps downwards. If the final step is upwards, then this can be thought of as changing s_{t+1} to $s_{t+1} + 1$. We encode this with $\varepsilon_t, \varepsilon'_t \in \{0, 1\}$, wherein the t -th linkage consists of $s_t + \varepsilon_t$ steps upwards followed by $\ell - s_t + \varepsilon'_t$ steps downwards. Since $\varepsilon'_t + \varepsilon_{t+1} = 1$, the sequence (ε'_t) is determined by (ε_t) , and $\sum \varepsilon_t + \varepsilon'_t = m$. Consequently,

$$\sum_t s_t + \varepsilon_t = \sum_t \ell - s_t + \varepsilon'_t = \frac{(\ell+1)m}{2}.$$

Setting $\mathcal{W}'_{\ell,m}$ as the set of walks enumerated by this trace moment method calculation,⁹ we have

$$\begin{aligned} |\mathcal{W}'_{\ell,m}| &\leq V(H) \cdot \sum_{s_1, \dots, s_m, \varepsilon_1, \dots, \varepsilon_m} \prod_{t=1}^m D^{\ell-s_t+\varepsilon'_t} \cdot \frac{\Delta}{D} \\ &= V(H) \cdot \sum_{s_1, \dots, s_m, \varepsilon_1, \dots, \varepsilon_m} D^{(\ell+1)m/2} \cdot \left(\frac{\Delta}{D}\right)^m \\ &\leq V(H) \cdot (\ell+1)^m \cdot 2^m \cdot D^{(\ell+1)m/2} \cdot \left(\frac{\Delta}{D}\right)^m. \end{aligned}$$

Again, taking $m \rightarrow \infty$ so $V(H)^{1/m} \rightarrow 1$, we get that

$$\|A^{(\ell)} A\| \leq D^{(\ell+1)/2} \cdot \frac{\Delta}{D} \cdot 2(\ell+1)$$

as claimed. ■

⁹ In this case, R is fixed to be $V(H)$.

As our final ingredient, we now prove a version of [Lemma 4.5](#) for trees with positive interactions and no external field.

Lemma 4.18. *Let H be a (Δ, D) -pseudorandom tree with respect to $S \subseteq V(H)$. Let A be the adjacency matrix of H . If J is a collection of interactions supported on the edges of H with values in $\left[0, \frac{\gamma}{\sqrt{D}}\right]$ for some $\gamma \in [0, 1)$, then*

$$\begin{aligned}\|\mathbf{Cov}(\mu_{J,0})_S\| &\leq \frac{1}{(1-\gamma)^2} \\ \|\mathbf{Cov}(\mu_{J,0})\| &\leq \frac{1}{(1-\gamma)^2} \cdot \frac{\Delta}{D} \\ \|\mathbf{Cov}(\mu_{J,0}) \cdot A\| &\leq \frac{1}{(1-\gamma)^2} \cdot \frac{2\Delta}{\sqrt{D}}.\end{aligned}$$

While only the first two of these bounds are needed in the current section, the third will come in handy in [Section 4.5](#) when we show an MLSI for the Ising model obtained after running stochastic localization, which is supported on a pseudorandom near-forest.

Proof. We start with the first two bounds, using the following explicit formula for the covariance matrix.

$$\mathbf{Cov}(\mu_{J,0}) = \sum_{\ell \geq 0} X^{(\ell)},$$

where $X_{i,j}^{(\ell)}$ is 0 if there is no length- ℓ path between i and j in H , and is otherwise equal to $\prod_{e \in P(i,j)} \tanh(J_e)$ where $P(i,j)$ is the unique such path. For $R \subseteq V(H)$, will use the bound:

$$\|\mathbf{Cov}(\mu_{J,0})_R\| \leq \sum_{\ell \geq 0} \|X_R^{(\ell)}\|. \quad (5)$$

Consider the matrix $Y^{(\ell)}$ obtained by replacing each nonzero entry of $X^{(\ell)}$ with $\left(\frac{\gamma}{\sqrt{D}}\right)^\ell$. Since the interactions are all nonnegative, $X^{(\ell)}$ is a nonnegative matrix, and $Y^{(\ell)}$ is entrywise at least $X^{(\ell)}$ since $\tanh(x) \leq x$ for $x \geq 0$. Consequently,

$$\|X_R^{(\ell)}\| \leq \|Y_R^{(\ell)}\|. \quad (6)$$

[Lemma 4.17](#) now implies that

$$\|Y_R^{(\ell)}\| \leq \left(\frac{\gamma}{\sqrt{D}}\right)^\ell \cdot D^{\ell/2} \cdot (\ell+1) \cdot \left(\frac{\Delta_R}{D}\right) = \gamma^\ell \cdot (\ell+1) \cdot \left(\frac{\Delta_R}{D}\right).$$

Therefore,

$$\|\mathbf{Cov}(\mu_{J,0})_R\| \leq \frac{\Delta_R}{D} \cdot \sum_{\ell \geq 0} (\ell+1)\gamma^\ell \leq \frac{1}{(1-\gamma)^2} \cdot \frac{\Delta_R}{D},$$

which recovers the desired bounds by plugging in $\Delta_R = \Delta$ for $R = V(H)$, and $\Delta_R = D$ for $R = S$.

For the third inequality, we again have

$$\mathbf{Cov}(\mu_{J,0}) \cdot A = \sum_{\ell \geq 0} \left\| X^{(\ell)} \cdot A \right\|.$$

Again, the non-negativity of $X^{(\ell)}$ and A implies that

$$\left\| X^{(\ell)} \cdot A \right\| \leq \left\| Y^{(\ell)} \cdot A \right\|.$$

The third bound in [Lemma 4.17](#) now implies that

$$\begin{aligned} \left\| Y^{(\ell)} \cdot A \right\| &\leq \left(\frac{\gamma}{\sqrt{D}} \right)^\ell D^{(\ell+1)/2} \cdot (\ell+1) \cdot 2 \cdot \left(\frac{\Delta}{D} \right) \\ &= \gamma^\ell \cdot (\ell+1) \cdot 2 \cdot \frac{\Delta}{\sqrt{D}}. \end{aligned}$$

Therefore,

$$\left\| \mathbf{Cov}(\mu_{J,0}) \cdot A \right\| \leq \sum_{\ell \geq 0} \gamma^\ell \cdot (\ell+1) \cdot 2 \cdot \frac{\Delta}{\sqrt{D}} \leq \frac{1}{(1-\gamma)^2} \cdot \frac{2\Delta}{\sqrt{D}},$$

completing the proof. ■

We are now ready to prove [Lemma 4.5](#).

Proof of [Lemma 4.5](#). The matrix $\mathbf{Cov}(\mu_{J,h})$ is a block diagonal matrix with a block per connected component of H , and so without loss of generality, it suffices to achieve a bound on the spectral norm in the case where H is a connected near-forest (a near-tree) — a tree with possibly one extra edge inserted. If H has tree-excess 0, the desideratum is trivial by [Fact 4.15](#) and [Lemmas 4.14](#) and [4.18](#). Hence, let us assume that H has tree-excess 1.

Let T be an arbitrary $((\Delta, D)$ -pseudorandom) spanning tree of H , such that the edge set of H consists of those in T along with an extra edge ij . By [Observation 4.12](#) and [Lemma 4.13](#), it suffices to consider the case where J has nonnegative interactions on all edges in T , and possibly a negative interaction on ij . Also let J' be the interaction matrix which is identical to J except that $J'_{ij} = J'_{ji} = 0$. Note that $|\sigma^\top J \sigma - \sigma^\top J' \sigma| \leq \frac{\gamma}{\sqrt{D}}$ for any $\sigma \in \{\pm 1\}^n$. Let R be equal to S or $V(H)$, with $\Delta_S = D$ and $\Delta_{V(H)} = \Delta$, and set v to be a unit vector that witnesses the operator norm of $\mathbf{Cov}(\mu_{J,h})_R$. Then,

$$\begin{aligned} \left\| \mathbf{Cov}(\mu_{J,h})_R \right\| &= v^\top \mathbf{Cov}(\mu_{J,h}) v \\ &= \mathbf{Var}_{\sigma \sim \mu_{J,h}}(\langle v, \sigma \rangle) && \text{(Observation 3.9)} \\ &\leq e^{2\gamma/\sqrt{D}} \cdot \mathbf{Var}_{\sigma \sim \mu_{J',h}}(\langle v, \sigma \rangle) && \text{(Lemma 3.8)} \\ &\leq e^{2\gamma/\sqrt{D}} \cdot \left\| \mathbf{Cov}(\mu_{J',h})_R \right\| && \text{(Observation 3.9)} \\ &\leq e^{2\gamma/\sqrt{D}} \cdot \left\| \mathbf{Cov}(\mu_{J',0})_R \right\| && \text{(Corollary 4.16)} \\ &\leq \frac{e^{2\gamma/\sqrt{D}}}{(1-\gamma)^2} \cdot \frac{\Delta_R}{D} && \text{(Lemma 4.18)} \end{aligned}$$

as desired. ■

4.4 Entropy conservation for near-forest decompositions

In this section, we shall bound the covariance matrix of the distribution obtained along the localization path for Ising models supported on a graph which admits a (Δ, D) -decomposition, towards satisfying the first requirement for applying [Theorem 3.20](#).

Let us now prove the main result of this section, which is restated here for convenience.

Theorem 4.6. *Let (B, H) be a (Δ, D) -decomposition of G , and let $J_B, J_H \in \mathbb{R}^{n \times n}$ be interaction matrices whose off-diagonal entries are supported on $E(B)$ and $E(H)$, respectively. Suppose that there are constants $\eta, K \in (0, 1)$ such that $\eta I_B \preceq J_B \preceq K I_B$, and that all interactions in J_H are in $\left[-\frac{\gamma}{\sqrt{D}}, \frac{\gamma}{\sqrt{D}}\right]$ for some $\gamma \in [0, 1)$. Finally, suppose that*

$$K \cdot \frac{e^{2\gamma/\sqrt{D}}}{(1-\gamma)^2} \leq 1 - \delta$$

for a positive constant $\delta > 0$. There exists a positive constant C not depending on Δ such that for any $0 \leq t \leq 1$ and external field $h \in \mathbb{R}^n$,

$$\mathbf{Cov}(\mu_{(1-t)J_B + tJ_H, h}) \preceq C(I_B + \Delta I_{H \setminus \partial H}).$$

Proof. Let $\rho > 0$ be a positive constant that we fix later. We use [Lemma 3.14](#), with the instantiation

$$M_1 = (1-t)J_B + \frac{t}{2}I_B + \frac{\rho}{\Delta}I_{H \setminus \partial H} \text{ and } M_2 = J_H.$$

Since $\eta I_B \preceq (1-t)J_B + \frac{t}{2}I_B \preceq K I_B$, we can reduce to the case where $t = 0$. Now, set $\delta' = \delta/2$. For $t = 0$ and an arbitrary external field h' , we have

$$\begin{aligned} M_1 \cdot \mathbf{Cov}(\mu_{M_2, h'}) \cdot M_1 &= \left(J_B + \frac{\rho}{\Delta}I_{H \setminus \partial H}\right) \cdot \mathbf{Cov}(\mu_{M_2, h'}) \cdot \left(J_B + \frac{\rho}{\Delta}I_{H \setminus \partial H}\right) \\ &\preceq (1 + \delta') \cdot J_B \cdot \mathbf{Cov}(\mu_{M_2, h'}) \cdot J_B + \left(1 + \frac{1}{\delta'}\right) \cdot \frac{\rho^2}{\Delta^2} I_{H \setminus \partial H} \cdot \mathbf{Cov}(\mu_{M_2, h'}) \cdot I_{H \setminus \partial H} \\ &= (1 + \delta') J_B \cdot \mathbf{Cov}(\mu_{M_2, h'}) \cdot J_B + \left(1 + \frac{1}{\delta'}\right) \cdot \frac{\rho^2}{\Delta^2} \mathbf{Cov}(\mu_{M_2, h'})_{H \setminus \partial H}. \end{aligned}$$

Above, the inequality follows from the observation that for symmetric matrices A_1, A_2 , a positive semidefinite matrix M , and any $\alpha > 0$,

$$(A_1 + A_2)M(A_1 + A_2) \preceq (1 + \alpha) A_1 M A_1 + \left(1 + \frac{1}{\alpha}\right) A_2 M A_2.$$

We now make the crucial observation that we can restrict the covariance to the bulk in the first term, that is,

$$J_B \cdot \mathbf{Cov}(\mu_{M_2, h'}) \cdot J_B = J_B \cdot \mathbf{Cov}(\mu_{M_2, h'})_B \cdot J_B.$$

Indeed, observe that $(J_B)_{ij}$ must be zero for any i, j where i or j lies in $H \setminus \partial H$, the interior of the near-forest. In particular, any such i or j is not in the bulk.

We are interested in showing the positive definiteness of

$$\left(J_B - (1 + \delta') \cdot J_B \cdot \mathbf{Cov}(\mu_{M_2, h'})_B \cdot J_B\right) + \frac{\rho}{\Delta} \left(I_{H \setminus \partial H} - \left(1 + \frac{1}{\delta'}\right) \cdot \frac{\rho}{\Delta} \cdot \mathbf{Cov}(\mu_{M_2, h'})_{H \setminus \partial H}\right).$$

Note that had we not added the extra $(\rho/\Delta)I_H$ when defining M_1 , the second term above would be absent, making the matrix entries zero on the rows and columns $V'(H) \setminus V'(B)$, which precludes a meaningful lower bound by a full-rank matrix.

By assumption, M_2 is supported on a (Δ, D) -pseudorandom near-forest, and the interaction strengths are in $\left[-\frac{\gamma}{\sqrt{D}}, \frac{\gamma}{\sqrt{D}}\right]$. Furthermore, since $\gamma < 1$, the first part of [Lemma 4.5](#) yields

$$\|\mathbf{Cov}(\mu_{M_2, h'})_B\| \leq \frac{e^{2\gamma/\sqrt{D}}}{(1-\gamma)^2} =: f(\gamma).$$

Consequently, [Fact 3.15](#) and the assumption on γ shows that

$$\begin{aligned} J_B - (1 + \delta') \cdot J_B \cdot \mathbf{Cov}(\mu_{M_2, h'}) \cdot J_B &\succeq \eta \cdot (1 - (1 + \delta') K f(\gamma)) I_B \\ &\succeq \eta \cdot \delta' \cdot I_B. \end{aligned}$$

Similarly, the second part of [Lemma 4.5](#) shows that

$$\|\mathbf{Cov}(\mu_{M_2, h'})_{H \setminus \partial H}\| \leq \|\mathbf{Cov}(\mu_{M_2, h'})_H\| \leq \frac{\Delta}{D} \cdot f(\gamma).$$

Selecting $\rho := \frac{D}{2f(\gamma) \cdot (1 + \frac{1}{\delta'})}$, we see that

$$I_{H \setminus \partial H} - \left(1 + \frac{1}{\delta'}\right) \cdot \frac{\rho}{\Delta} \cdot \mathbf{Cov}(\mu_{M_2, h'})_{H \setminus \partial H} \succeq \frac{1}{2} I_{H \setminus \partial H}.$$

Thus, we have that

$$\begin{aligned} M_1 - M_1 \cdot \mathbf{Cov}(\mu_{M_2, h'}) \cdot M_1 &\succeq \eta \delta' \cdot I_B + \frac{\rho}{2\Delta} \cdot I_{H \setminus \partial H} \\ &\succeq c \left(I_B + \frac{1}{\Delta} I_{H \setminus \partial H} \right), \end{aligned}$$

for some universal constant c (that does not depend on Δ). The claim follows by applying [Lemma 3.14](#) and setting $C = c^{-1}$. ■

4.5 MLSI for Ising models on pseudorandom near-forests

Our goal in this section will be to prove a modified log-Sobolev inequality for the Ising model obtained after stochastic localization of the original Ising model, which we recall for convenience.

Theorem 4.8. *Let H be a (Δ, D) -pseudorandom near-forest, and $J \in \mathbb{R}^{n \times n}$ an interaction matrix supported on H with all interactions in $\left[-\frac{\gamma}{\sqrt{D}}, \frac{\gamma}{\sqrt{D}}\right]$ for some $\gamma \in [0, 1)$. There is a constant c (depending on γ) such that for any external field $h \in \mathbb{R}^n$,*

$$C_{\text{MLSI}}(\mu_{J, h}) \geq \frac{1}{n e^{c\Delta}}.$$

Fact 3.7 implies that it suffices to prove an MLSI of $\frac{1}{|C|e^\Delta}$ for each component C of this near-forest, a “near-tree”. By a simple comparison argument, it in fact suffices to prove an MLSI for pseudorandom trees with ferromagnetic interactions — we shall elaborate on this in the proof.

Lemma 4.19. *Let H be a (Δ, D) -pseudorandom tree on n vertices, and J an interaction matrix supported on H with all interactions in $\left[0, \frac{\gamma}{\sqrt{D}}\right]$ for some $\gamma \in [0, 1)$. There is a constant c (depending on γ) such that for any external field $h \in \mathbb{R}^n$,*

$$C_{\text{MLSI}}(\mu_{J,h}) \geq \frac{1}{nc^{c\Delta}}.$$

Remark 4.20. In fact, our proof works in a slightly broader range of the interactions, only requiring that for some $\gamma \in [0, 1)$, all the interactions are bounded (in absolute value) by $\tanh^{-1}\left(\frac{\gamma}{\sqrt{D}}\right)$. We conjecture that this is optimal.

In this section, we only provide a proof of the above in the setting where all the nonzero interactions are exactly equal to $\frac{\gamma}{\sqrt{D}}A$. The more general proof is provided in [Appendix A](#).

Proof. Let H be a tree as in the lemma statement, A its adjacency matrix, and $J = \frac{\gamma}{\sqrt{D}}$ the corresponding interaction matrix. Also set D_H be the (diagonal) degree matrix of H .

To prove this, we shall once again use [Lemma 3.18](#). The control matrix will be of the form $C = (J + E)^{1/2} \succeq 0$ for a non-negative diagonal matrix E . Due to the ferromagneticity of the system, C^2 has all non-negative entries. At time $0 \leq t \leq 1$, J_t will have the form $(1-t)J - tE$ — in particular, at time $t = 1$, μ_t is almost surely a product distribution so $C_{\text{MLSI}}(\mu_t) = \Omega(1/n)$. In the context of [Lemma 3.18](#), our goal is to bound, for all external fields $h \in \mathbb{R}^n$, the operator norm

$$\left\| C \cdot \mathbf{Cov}(\mu_{(1-t)J,h}) \cdot C \right\| \leq \left\| \mathbf{Cov}(\mu_{(1-t)J,h}) \cdot C^2 \right\|. \quad (7)$$

By [Corollary 4.16](#) (using the matrix C^2), it suffices to bound the above for $h = 0$. For $h = 0$, we recall that if $A^{(\ell)}$ is the matrix where the uv -th entry is 1 if u and v are distance exactly ℓ apart and 0 otherwise,

$$\mathbf{Cov}(\mu_{(1-t)J,0}) \leq \sum_{\ell \geq 0} \left(\frac{\gamma}{\sqrt{D}} \right)^\ell A^{(\ell)}$$

in the sense that the matrix on the right is entry-wise at least the matrix on the left. In particular,

$$\left\| \mathbf{Cov}(\mu_{(1-t)J,h}) \cdot C^2 \right\| \leq \left\| \sum_{\ell \geq 0} \left(\frac{\gamma}{\sqrt{D}} \right)^\ell A^{(\ell)} \cdot C^2 \right\|. \quad (8)$$

For ease of notation, denote $s_0 = \frac{\gamma}{\sqrt{D}}$. Recalling [Definition 3.22](#), set

$$\text{BH}(s) = \frac{(D_H - I)s^2 - As + I}{1 - s^2},$$

so by [Fact 3.25](#),

$$\text{BH}(s_0)^{-1} = \sum_{\ell \geq 0} \left(\frac{\gamma}{\sqrt{D}} \right)^\ell A^{(\ell)}.$$

We choose

$$C = ((1 - s_0^2) \cdot \text{BH}(-s_0))^{1/2}. \quad (9)$$

Now,

$$\begin{aligned} \sum_{\ell \geq 0} \left(\frac{\gamma}{\sqrt{D}} \right)^\ell A^{(\ell)} \cdot C^2 &= (1 - s_0^2) \cdot \text{BH}(s_0)^{-1} \cdot \text{BH}(-s_0) \\ &= (1 - s_0^2) \cdot \text{BH}(s_0)^{-1} \cdot \left(\text{BH}(s_0) + \frac{2s_0}{1 - s_0^2} \cdot A \right) \\ &= (1 - s_0^2) \cdot I + 2s_0 \cdot \text{BH}(s_0)^{-1} \cdot A. \end{aligned}$$

Using the third bound in [Lemma 4.18](#) with (7) and (8), we get the desired claim that

$$\begin{aligned} \left\| C \cdot \mathbf{Cov} \left(\mu_{(1-t)J,h} \right) \cdot C \right\| &\leq \left\| (1 - s_0^2) \cdot I + 2s_0 \cdot \text{BH}(s_0)^{-1} \cdot A \right\| \\ &\leq 1 + \frac{8}{(1 - \gamma)^2} \cdot \frac{\Delta}{D}. \end{aligned}$$

Given that the product measure obtained at time 1 satisfies an MLSI with constant $1/n$, it follows by [Lemma 3.18](#) and [Theorem 3.19](#) that

$$\begin{aligned} C_{\text{MLSI}}(\mu_{J,h}) &\geq \frac{1}{n} \cdot \exp \left(- \int_0^1 \sup_{h \in \mathbb{R}^n} \left\| C \cdot \mathbf{Cov} \left(\mu_{(1-t)J,h} \right) \cdot C \right\| dt \right) \\ &\geq \frac{1}{n} \cdot \exp \left(- \left(1 + \frac{8}{(1 - \gamma)^2} \cdot \frac{\Delta}{D} \right) \right) = \frac{1}{ne^{O(\Delta)}} \end{aligned}$$

as desired. ■

Before moving on, we make a handful of remarks about the above proof, specifically the choice of control matrix in (9):

- The choice is well-defined due to [Corollary 3.24](#), which implies that $\text{BH}(-s_0) \succ 0$.
- It is important that we provide a negative argument to BH so that C^2 is equal to some positive scaling of the interaction matrix (which in turn is a scaling of the adjacency matrix because all the nonzero interactions are equal) plus a diagonal matrix – this is required for stochastic localization to eventually kill the interaction matrix. Note that as a consequence of this, C^2 has all non-negative entries as desired.
- It is also important that the precise negative argument we provide is equal to $-s_0$ in order to be able to express C^2 as $\text{BH}(s_0)$ plus some scaling of the adjacency matrix – otherwise, there would also be a term corresponding to the degree matrix in the definition of the Bethe Hessian. This would cause issues in the methods we use to bound the operator norm.

The above lemma easily yields a similar result even when interactions are allowed to be negative, as long as their absolute values satisfy the same bound.

Corollary 4.21. *Let H be a (Δ, D) -pseudorandom tree on n vertices, and J an interaction matrix supported on H with all interactions in $\left[-\frac{\gamma}{\sqrt{D}}, \frac{\gamma}{\sqrt{D}}\right]$. For any external field h ,*

$$C_{\text{MLSI}}(\mu_{J,h}) \geq \frac{1}{ne^{O(\Delta)}}.$$

Proof. This immediately follows using [Lemma 4.19](#) in conjunction with [Observation 4.12](#) and [Lemma 4.13](#), since H is a tree. ■

Given the above, [Theorem 4.8](#) is near-immediate using the Holley-Stroock perturbation principle [Lemma 3.8](#).

Proof of Theorem 4.8. As remarked earlier, by [Fact 3.7](#), it suffices to prove an MLSI of $\Omega\left(\frac{1}{|C|e^\Delta}\right)$ for each component C of H , which is a near-tree. Let μ be the corresponding Ising model on this component, with interaction matrix $J|_C$ and external field $h|_C$. Let T be an arbitrary spanning tree of C (which must also be (Δ, D) -pseudorandom), and ν the Ising model supported on T with the same interaction strengths as in μ . Observe that because C has at most one edge in addition to those in T , and the interaction strengths are bounded by $\frac{\gamma}{\sqrt{D}}$ in absolute value, we have

$$e^{-\frac{2\gamma}{\sqrt{D}}} \leq \frac{\mu(\sigma)}{\nu(\sigma)} \leq e^{\frac{2\gamma}{\sqrt{D}}}.$$

Furthermore, ν satisfies an MLSI by [Corollary 4.21](#), so [Lemma 3.8](#) completes the proof. ■

5 Glauber dynamics for centered stochastic block models

In this section, we establish the following result proving an MLSI for Ising models where the interactions are proportional to the centered adjacency matrix \bar{A}_G for a sparse random graph G .

Theorem 5.1. *Fix constants $d > 1, \lambda \geq 0$. There exists a universal constant $\beta > 0$ such that the following is true. Let $\bar{A}_G := A_G - \mathbf{E}[A_G|\sigma]$ be the centered adjacency matrix of $G \sim \text{SBM}(n, d, \lambda)$. With high probability over the draw of G , for any external field h , $C_{\text{MLSI}}(\mu_{(\beta/\sqrt{d})\bar{A}_G,h}) \geq n^{-1-o_d(1)}$.*

Given this, [Fact 3.6](#) immediately yields [Theorem 1.3](#).

Due to the presence of the weak (but nonzero) centering interactions, the analysis turns out to be significantly more delicate than that in [Section 4](#). To prove an MLSI, one still uses [Theorem 3.20](#), where the initial localization scheme anneals away the bulk interactions. However, the annealed Ising model now has small but nonzero interactions that are supported beyond the near forest H .

Thus, we need to prove analogous versions of [Theorems 4.6](#) and [4.8](#) for entropy conservation and an MLSI for the annealed model. We state the results below and prove them in subsequent sections.

Theorem 5.2. *Let $G \sim \text{SBM}(n, d, \lambda)$ for any constant $d \geq 1$, and let $B = G_1$ and $H = G_2$ be the bulk component and near-forest component respectively that are guaranteed by [Theorem 4.9](#). There exists a constant $\beta^* > 0$ such that with high probability over the draw of G and any external field h , the following*

holds for all $\beta < \beta^*$. Set $J_B = \frac{\beta}{\sqrt{d}}(A_{G_1} - \mathbf{E}[A|\sigma]_{G_1})$ and $J_R = \frac{\beta}{\sqrt{d}}\bar{A}_G - J_B$. There exists a positive constant C such that for any $0 \leq t \leq 1$ and external field $h \in \mathbb{R}^n$,

$$\mathbf{Cov}(\mu_{(1-t)J_B+J_R,h}) \preceq C(I_B + \Delta I_{H \setminus \partial H}).$$

Theorem 5.3. Let $J_R \in \mathbb{R}^{n \times n}$ be the interaction matrix for the perturbed near-forest as defined in [Theorem 5.2](#). Then, for any external field h ,

$$C_{\text{MLSI}}(\mu_{J_R,h}) \geq \frac{1}{n^{1+o_d(1)}}.$$

Given the above lemmas, the proof of [Theorem 5.1](#) is exactly like that of [Theorem 4.1](#).

Proof of [Theorem 5.1](#). By [Theorem 5.2](#), for sufficiently low inverse temperature β , it holds with high probability that all $0 \leq t \leq 1$ and external fields $h \in \mathbb{R}^n$,

$$\mathbf{Cov}(\mu_{(1-t)J_B+J_R,h}) \preceq C(I_B + o(\log n) \cdot I_{H \setminus \partial H}).$$

With $\tilde{J}_B = J_B + 2\beta(1 + \varepsilon)I_B$, [Lemma 3.29](#) implies that

$$0 \preceq \tilde{J}_B \preceq 4\beta(1 + \varepsilon)I_B.$$

Thus,

$$\exp\left(-\left\|\tilde{J}_B^{1/2} \cdot \mathbf{Cov}\left(\mu_{(1-t)J_B+J_R,h}\right) \cdot \tilde{J}_B^{1/2}\right\|\right) = \Omega(1).$$

On the other hand, [Theorem 5.3](#) says that for all h ,

$$C_{\text{MLSI}}(\mu_{J_R,h}) \geq \frac{1}{n^{1+o_d(1)}}.$$

[Theorem 3.20](#) completes the proof. ■

5.1 Technical overview for centered stochastic block models

We now elucidate the high-level strategy for proving [Theorems 5.2](#) and [5.3](#). The annealed Ising model is $\mu_{J_R,h}$ for some external field h ; here J_R stands for the remainder of the interactions outside of the bulk. We will refer to J_R as the *perturbed near-forest* interactions, and $\mu_{J_R,h}$ as the perturbed near-forest Ising model. Let us introduce some notation to separate out the different pieces of J_R .

Let $H' = V'(H)$ denote the set of non-isolated vertices in H , and let $B' := V(B) \setminus H'$ denote the set of vertices on the interior of the bulk (we slightly abuse notation here for notational simplicity). Furthermore, for any two matrices $M, N \in \mathbb{R}^{n \times n}$, the Hadamard product $M \circ N \in \mathbb{R}^{n \times n}$ denotes the entrywise product of M and N .

We can decompose $J_R = M_H + E_H + E$, with the latter defined explicitly as

$$J_R = \underbrace{\frac{\beta}{\sqrt{d}}(\bar{A}_H \circ A_H)}_{M_H} - \underbrace{\frac{\beta}{\sqrt{d}}\mathbf{E}[A|\sigma]_H \circ (\mathbf{1}_{H'}\mathbf{1}_{H'}^\top - A_H)}_{E_H} - \underbrace{\frac{\beta}{\sqrt{d}}(\mathbf{E}[A|\sigma]_{H',B'} + \mathbf{E}[A|\sigma]_{B',H'})}_{E}, \quad (10)$$

where

$$\mathbf{E}[A|\sigma]_{H',B'} = \frac{d}{n} \left(\mathbf{1}_{H'}\mathbf{1}_{B'}^\top + \frac{\lambda}{\sqrt{d}}\sigma_{H'}\sigma_{B'}^\top \right) = \mathbf{E}[A|\sigma]_{B',H'}^\top.$$

When G is a sparse Erdős-Rényi graph (instead of the more general stochastic block model),

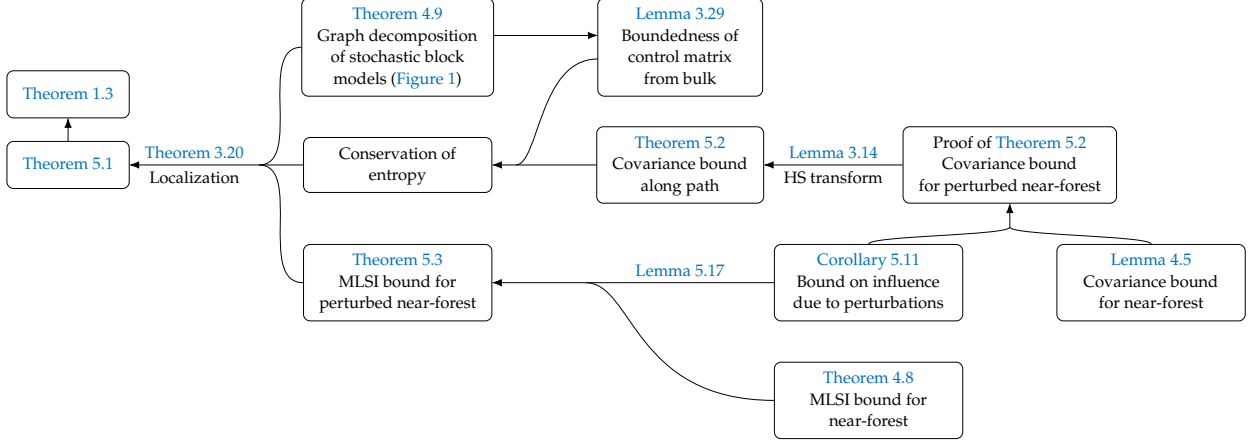


Figure 3: A flow chart outlining our proof of an MLSI for the centered adjacency matrix. The main difference in the structure of the proof compared to Figure 2 are the three boxes on the bottom right, which handle the interactions present on nonedges of the graph due to centering.

- $\frac{\sqrt{d}}{\beta} \cdot M_H$ is equal to $\left(1 - \frac{d}{n}\right)$ on the edges within the near-forest H .
- $\frac{\sqrt{d}}{\beta} \cdot E_H$ is equal to $\frac{d}{n}$ on all the non-edges in the near-forest H .
- $\frac{\sqrt{d}}{\beta} \cdot E$ is equal to $\frac{d}{n}$ on all pairs between the near-forest H and the interior of the bulk B' .

In particular, J_R is equal to zero on the $B' \times B'$ pairs.

Note that if only M_H were present, then we would be back in the setting of Lemma 4.5. Accordingly, we shall show that the perturbation terms E and E_H do not affect the covariances or MLSIs by too much.

Crucially, observe that the perturbation interactions are $O\left(\frac{d}{n}\right)$, and nonzero if and only if the edge is incident to a vertex of H . In particular, for any isolated vertex, corresponding to a vertex in B' , the perturbation vanishes. Moreover, we show that the size of each connected component is $o(n)$, which means that the perturbation interactions are genuinely small relative to the scale of connected components. This motivates introducing a Markov chain whose states correspond to the nontrivial connected components C_i of H with size strictly greater than 1.

Since the perturbation interactions are small, we can show that the spectral norm version of Dobrushin's condition [Hay06] is satisfied for this Markov chain. In the end, this allows us to pass to the simpler Ising model supported on a single connected component C_i , where we can apply our results from Section 4.1.

As in the technical overview, we outline the structure of our proof in Figure 3. The new elements in the proof compared to the simpler spin glass setting are on the bottom right of the chart — a bound on a certain Dobrushin influence matrix is the primary new tool that is required.

To rigorously execute the above plan, we will need some additional machinery. We introduce these tools in Section 5.3 to prove Theorem 5.3. Using the same tools, we then prove Theorem 5.2 in Section 5.4.

5.2 A crash course on spectral independence

We first introduce some additional tools that we use to establish the MLSI for the perturbed near-forest. These ingredients are stated somewhat abstractly, but we will next point out what the definitions correspond to in our application and tie it all together when we apply them to sparse random graphs.

Influence matrices and spectral independence. We work in the following general setup. Let ν be a distribution on a generic product domain $\mathcal{P} = \mathcal{P}_1 \times \cdots \times \mathcal{P}_k$. A key object of interest associated to ν is its *Dobrushin influence matrix*, which is very standard in the literature on mixing times of Glauber dynamics.

Definition 5.4 (Dobrushin Influence matrix [Dob68, BD97]). The *Dobrushin influence matrix* of ν , denoted \mathcal{R}_ν , is a $k \times k$ matrix where:

$$\mathcal{R}_\nu[i, j] := \max_{\substack{\sigma_{-i}, \tau_{-i} \\ \text{that agree on coordinates} \\ [k] \setminus \{i, j\}}} d_{\text{TV}}(\nu | \sigma_{-i}, \nu | \tau_{-i}).$$

The following was proved in [AJK⁺24], with a slightly weaker version presented in [Liu21, BCC⁺22].

Theorem 5.5 ([AJK⁺24, Theorem 43]). *If $\|\mathcal{R}_\nu\| \leq \delta < 1$, then ν is $(\frac{1}{1-\delta})$ -spectrally independent.*

In order to prove the MLSI for the perturbed Ising model, we will need to show that the block Glauber dynamics on connected components introduced earlier has entropy factorization. To this end, we will need the notion of a *tilt*.

Definition 5.6 (Tilts). A *tilt* $\lambda = (\lambda_1, \dots, \lambda_k)$ is a collection of functions $\lambda_i : \mathcal{P}_i \rightarrow \mathbb{R}_{>0}$ for $i \in [k]$, and the *tilted distribution* $\nu * \lambda$ is:

$$(\nu * \lambda)(x) \propto \nu(x) \cdot \prod_{i=1}^k \lambda_i(x_i).$$

The following is a consequence of [AJK⁺21a, Theorem 4 and Theorem 5] and [AASV21, Proposition 20 and Remark 70].

Lemma 5.7. *Suppose for every tilt λ , the distribution $(\nu * \lambda)$ is γ -spectrally independent for some constant $\gamma > 0$. Then ν satisfies the following conservation of entropy property. For any function $f : \mathcal{P} \rightarrow \mathbb{R}_{>0}$,*

$$\mathbf{E}_{i \sim [k]} \mathbf{E}_{x_{-i} \sim \nu_{-i}} \text{Ent}_{\nu_i}[f] \geq \frac{1}{k^\gamma} \cdot \text{Ent}_\nu[f],$$

where ν_i is the conditional marginal of site i conditioned on all other sites.

The following is an immediate corollary of [Theorem 5.5](#) and [Lemma 5.7](#).

Corollary 5.8. *Suppose for every tilt λ , the influence matrix satisfies:*

$$\|\mathcal{R}_{\nu * \lambda}\| \leq \delta < 1.$$

Then for every function $f : \mathcal{P} \rightarrow \mathbb{R}_{>0}$,

$$\mathbf{E}_{i \sim [k]} \mathbf{E}_{x_{-i} \sim \nu_{-i}} \text{Ent}_{\nu_i}[f] \geq \frac{1}{k^{1-\delta}} \cdot \text{Ent}_\nu[f].$$

5.3 MLSI for Ising models on perturbed near-forests

We now establish an MLSI for perturbed Ising models which satisfy certain assumptions ([Assumption 5.9](#)). Following this, we verify these assumptions for our concrete setting ([Lemma 5.17](#)). For convenience, we restate our goal.

Theorem 5.3. *Let $J_R \in \mathbb{R}^{n \times n}$ be the interaction matrix for the perturbed near-forest as defined in [Theorem 5.2](#). Then, for any external field h ,*

$$C_{\text{MLSI}}(\mu_{J_R, h}) \geq \frac{1}{n^{1+o_d(1)}}.$$

Many of the tools introduced in this subsection are reused in the next subsection, where we bound the covariance of the Ising models obtained along the localization path.

MLSI for perturbed Ising models. Consider a collection of connected components C_1, \dots, C_m , as well as a collection of isolated vertices $v_1, \dots, v_{m'}$, and set $n = \sum_i |C_i| + m'$. We study a “base” Ising model with interaction matrix $J \in \mathbb{R}^{n \times n}$ supported on the edges of the (C_i) and diagonal, along with an arbitrary external field $h \in \mathbb{R}^n$.

A key assumption we make is the following.

Assumption 5.9. For each C_i , and for any external field h , the Ising model $\mu_{J_{C_i}, h}$ on $\{\pm 1\}^{V(C_i)}$ satisfies an MLSI with constant ζ . Additionally, suppose that $\eta, \delta, \theta > 0$ are small enough parameters (which can be shrinking with n) such that:

$$\begin{aligned} \max_i |C_i| &\leq n^\theta \\ \eta \cdot \sum_i |C_i|^2 &< \delta \\ 2 \cdot \eta^2 \cdot m' \cdot \sum_i |C_i|^2 &< \delta. \end{aligned}$$

We will be interested in perturbed Ising models of the form $J + E$, where every entry of E is at most η , and $E_{ij} = 0$ if both i and j are among the $(v_k)_{1 \leq k \leq m'}$.

To keep notation succinct, we use μ as shorthand for the Gibbs distribution of the perturbed Ising model $\mu_{J+E, h}$.

In the setting of [Corollary 5.8](#), each \mathcal{P}_i corresponds to the collection of spins on the vertices of one of the connected components C_j or the spin on one of the isolated vertices v_j . We index our influence matrix by $C_1, \dots, C_m, v_1, \dots, v_{m'}$. The following is a consequence of a straightforward calculation.

Observation 5.10. For any tilt λ , the influence matrix satisfies:

$$\begin{aligned} \mathcal{R}_{\mu * \lambda}[C_i, C_j] &\leq 1 - \exp(-\eta |C_i| \cdot |C_j|) \\ \mathcal{R}_{\mu * \lambda}[C_i, v_j] &\leq 1 - \exp(-\eta |C_i|) \\ \mathcal{R}_{\mu * \lambda}[v_i, C_j] &\leq 1 - \exp(-\eta |C_j|) \\ \mathcal{R}_{\mu * \lambda}[v_i, v_j] &= 0. \end{aligned}$$

Using $\exp(x) \geq 1 + x$, we get the following as a corollary.

Corollary 5.11. *For any tilt λ ,*

$$\|\mathcal{R}_{\mu*\lambda}\|_F \leq \sqrt{\eta^2 \cdot \left(\sum_i |C_i|^2\right)^2 + 2 \cdot \eta^2 \cdot m' \cdot \sum_i |C_i|^2}.$$

Hence, under [Assumption 5.9](#), $\|\mathcal{R}_{\mu*\lambda}\|_{\text{op}} \leq \delta + \sqrt{\delta} \leq 2\sqrt{\delta}$.

Thus, by [Corollary 5.8](#), we get the following.

Corollary 5.12. *Let $k = m + m'$. Under [Assumption 5.9](#), the perturbed Ising model μ satisfies the following entropy conservation bound.*

$$\mathbf{E}_{b \sim \{C_1, \dots, C_m\} \cup \{v_1, \dots, v_{m'}\}} \mathbf{E}_{x \sim \mu_{-b}} \text{Ent}_{\mu_b}[f] \geq \frac{1}{k^{1-2\sqrt{\delta}}} \cdot \text{Ent}_{\mu}[f].$$

Next, we observe that based on the assumptions on the parameters, for any individual connected component C_i , any Ising model with the perturbed interaction matrix restricted to C_i satisfies an MLSI.

Lemma 5.13. *Under [Assumption 5.9](#), for any external field h , we have*

$$\exp\left(-O\left(\eta \cdot n^{2\theta}\right)\right) \leq \frac{\mu_{(J+E)_{C_i}, h}(x)}{\mu_{J_{C_i}, h}(x)} \leq \exp\left(O\left(\eta \cdot n^{2\theta}\right)\right).$$

Consequently, the Glauber dynamics on $\mu_{(J+E)_{C_i}, h}$ satisfies

$$C_{\text{MLSI}}(\mu_{(J+E)_{C_i}, h}) \geq \exp\left(-O\left(\eta \cdot n^{2\theta}\right)\right) \cdot \zeta.$$

Proof. For any $x \in \{\pm 1\}^{V(C_i)}$, the first bound easily follows from the bound on $|C_i|$ and the entries of E . The second statement then follows from [Lemma 3.8](#). \blacksquare

We are now ready to prove the MLSI for Glauber dynamics for μ . Recall that $n = \sum_i |C_i| + m'$, so μ is a distribution on $\{\pm 1\}^n$.

Lemma 5.14. *Under [Assumption 5.9](#), Glauber dynamics on μ satisfies*

$$C_{\text{MLSI}}(\mu) \geq \exp\left(-O\left(\eta \cdot n^{2\theta}\right)\right) \cdot \frac{\zeta}{n^{1-2\sqrt{\delta}}} \cdot \left(1 - \frac{\delta}{\eta n}\right).$$

Proof. We can sample a uniform index i from $[n]$ by the following two-step process. First, sample b from a certain distribution \mathcal{D} on $\{C_1, \dots, C_m\} \cup \{v_1, \dots, v_{m'}\}$, and then choose i as a uniformly random vertex from b . In particular, the distribution \mathcal{D} is defined as:

$$\Pr[b = v_i] = \frac{1}{n} \quad \Pr[b = C_i] = \frac{|C_i|}{n}.$$

For any b , and any pinning x_{-b} to the spins outside of b , the resulting conditional distribution on x_b is an Ising model with interaction matrix $(J + E)_b$ and some external field $h(x_{-b})$ depending on the pinning. This lets us write the distribution μ as a mixture:

$$\mu = \mathbf{E}_{b \sim \mathcal{D}} \mathbf{E}_{x_{-b} \sim \mu_{-b}} \mu^{(J+E)_b, h(x_{-b})}.$$

We can succinctly express the mixture distribution as ρ , the pair (b, x_{-b}) as z , and the resulting Ising model as π_z . To prove the MLSI, we write the following chain of inequalities. For any function $f : \{\pm 1\}^n \rightarrow \mathbb{R}_{\geq 0}$:

$$\begin{aligned} \mathcal{E}_\mu(f, \log f) &\geq \mathbf{E}_{z \sim \rho} \mathcal{E}_{\pi_z}(f, \log f) \\ &\geq \exp\left(-O\left(\eta \cdot n^{2\theta}\right)\right) \cdot \zeta \cdot \mathbf{E}_{z \sim \rho} \text{Ent}_{\pi_z}[f] && \text{(MSI for } \pi_z, \text{ Lemma 5.13)} \\ &= \exp\left(-O\left(\eta \cdot n^{2\theta}\right)\right) \cdot \zeta \cdot \mathbf{E}_{b \sim \mathcal{D}} \mathbf{E}_{x_{-b} \sim \mu_{-b}} \text{Ent}_{\pi_{(b, x_{-b})}}[f] \\ &\geq \exp\left(-O\left(\eta \cdot n^{2\theta}\right)\right) \cdot \zeta \cdot \left(1 - \frac{\delta}{\eta n}\right) \cdot \mathbf{E}_{b \sim \{C_1, \dots, C_m\} \cup \{v_1, \dots, v_{m'}\}} \mathbf{E}_{x_{-b} \sim \mu_{-b}} \text{Ent}_{\pi_{(b, x_{-b})}}[f] \\ &\geq \exp\left(-O\left(\eta \cdot n^{2\theta}\right)\right) \cdot \frac{\zeta}{n^{1-2\sqrt{\delta}}} \cdot \left(1 - \frac{\delta}{\eta n}\right) \cdot \text{Ent}_\mu[f], && \text{(Corollary 5.12)} \end{aligned}$$

where in the fourth line we move from \mathcal{D} to the uniform distribution over the parts using the fact that for any component C_i ,

$$\frac{|C_i|/n}{1/(m+m')} \geq \frac{m}{n} \geq \frac{n - \sum_{i:|C_i|>1} |C_i|}{n} \geq 1 - \frac{\delta}{\eta n}.$$

This completes the proof. ■

Component sizes of sparse random graphs. To prove [Theorem 5.3](#) and obtain explicit parameter dependencies, we will need to verify [Assumption 5.9](#) for sparse random graphs. The following results give finer control on the near-forest decomposition; we defer their proofs to [Section 6](#).

Lemma 5.15. *Let $G \sim \text{SBM}(n, d, \lambda)$ for any constant $d \geq 1$. Let G_2 be the near-forest component guaranteed by [Theorem 4.9](#), and write G_2 as the union of connected components $G_2 = \bigsqcup_i C_i$.*

For $\varepsilon \geq \Omega(\log d/d)^{1/3}$, the following holds with $\alpha := \exp(-c\varepsilon^3 d)$ for some absolute constant $c > 0$. With probability $1 - o(1)$ over G ,

$$\sum_{i:|C_i|>1} |C_i|^2 \leq \alpha n.$$

Lemma 5.16. *With probability $1 - o(1)$, the maximum size of a connected component is at most $n \frac{c \log(d(1+\varepsilon))}{\varepsilon^3 d}$ for some absolute constant $c > 0$.*

Let us put everything together and prove that [Assumption 5.9](#) holds for sparse random graphs; this will also be required later when we bound the covariance of the Ising models obtained along the localization path.

Lemma 5.17. Let $G \sim \text{SBM}(n, d, \lambda)$. Then μ_{J_R} satisfies [Assumption 5.9](#) with $m' \leq n$ and

$$(\eta, \delta, \theta, \zeta) \leftarrow \left(\frac{2d}{n}, 8\alpha d^2, \frac{c \log(d(1+\varepsilon))}{\varepsilon^3 d}, n^{-o_d(1)} \right),$$

where c is a positive universal constant. Consequently, for $\varepsilon = \Omega(\log d/d)^{1/3}$, Glauber dynamics on μ_{J_R} satisfies

$$C_{\text{MLSI}}(\mu_{J_R}) \geq \frac{1}{n^{1+o_d(1)}}.$$

Proof. Let $\text{cc}(H)$ denote the connected components of H , and write $H = \sqcup_{C \in \text{cc}(H)} C$. Let $\mathcal{S} = \{C \in \text{cc}(H) : |C| > 1\}$ be the nontrivial connected components, and $\mathcal{T} = \{C \in \text{cc}(H) : |C| = 1\}$ be the isolated vertices. In our setup, the error $E_H + E$ has entries whose magnitude are bounded by $\frac{d}{n} \left(1 + \frac{|\lambda|}{\sqrt{d}}\right) \leq \frac{2d}{n}$. On the other hand, [Lemma 5.15](#) implies that $|\mathcal{S}| \leq \alpha n$, where $\alpha = \exp(-c\varepsilon^3 d)$, and we trivially have $|\mathcal{T}| \leq n$.

Hence, we will verify the conditions for $\eta = \frac{2d}{n}$ and $m' = n$. Plugging these parameters in, we need δ to satisfy

$$\begin{aligned} \frac{2d}{n} \cdot \sum_{C \in \mathcal{S}} |C|^2 &< \delta \\ 2 \cdot \left(\frac{2d}{n}\right)^2 \cdot n \cdot \sum_{C \in \mathcal{S}} |C|^2 &< \delta. \end{aligned}$$

Applying [Lemma 5.15](#) again, we see that the above inequalities are satisfied for $\delta > 8\alpha d^2$. Hence, we pick $\delta = 8\alpha d^2$, and we easily compute $1 - \frac{\delta}{\eta n} = 1 - 4\alpha d = 1 - o_d(1)$.

Turning now to θ , we need θ to satisfy

$$\max_{C \in \mathcal{S}} |C| \leq n^\theta.$$

[Lemma 5.16](#) implies that we can take $\theta = \frac{c \log(d(1+\varepsilon))}{\varepsilon^3 d} = o_d(1)$ for some positive constant c . Next, by the proof of [Theorem 4.8](#), we can take $\zeta = n^{-\theta - o_n(1)} = n^{-o_d(1)}$ (observe that we get $n^{-\theta - o(1)}$ instead of $n^{-1 - o(1)}$ here because we are considering the Ising model over $\{\pm 1\}^{V(C_i)}$ instead of $\{\pm 1\}^n$).

Finally, plugging these parameters into [Lemma 5.14](#), we have

$$\exp(-O(\eta \cdot n^{2\theta})) = 1 - O\left(\frac{d}{n^{1-2\theta}}\right)$$

By picking the constant factor in ε large enough, we can ensure that θ is sufficiently small so that the above term is $1 - o_n(1)$. ■

Clearly, [Lemma 5.17](#) directly proves [Theorem 5.3](#).

5.4 Entropy conservation for centered stochastic block models

In this section, we prove [Theorem 5.2](#), which is used to establish entropy conservation for centered Ising models on sparse SBMs.

Theorem 5.2. *Let $\mathbf{G} \sim \text{SBM}(n, d, \lambda)$ for any constant $d \geq 1$, and let $B = \mathbf{G}_1$ and $H = \mathbf{G}_2$ be the bulk component and near-forest component respectively that are guaranteed by [Theorem 4.9](#). There exists a constant $\beta^* > 0$ such that with high probability over the draw of \mathbf{G} and any external field h , the following holds for all $\beta < \beta^*$. Set $J_B = \frac{\beta}{\sqrt{d}}(A_{G_1} - \mathbf{E}[A|\sigma]_{G_1})$ and $J_R = \frac{\beta}{\sqrt{d}}\overline{A}_G - J_B$. There exists a positive constant C such that for any $0 \leq t \leq 1$ and external field $h \in \mathbb{R}^n$,*

$$\mathbf{Cov}(\mu_{(1-t)J_B+J_R,h}) \preceq C(I_B + \Delta I_{H \setminus \partial H}).$$

Let B, H be a (Δ, D) -decomposition as guaranteed by [Theorem 4.9](#). As before, we will use [Lemma 3.14](#) with

$$M_1 = (1-t)J_B + \theta I_B + \frac{\rho}{\Delta} I_{H \setminus \partial H} \quad \text{and} \quad M_2 = J_R,$$

where we can take $\theta = \frac{t}{2} + \eta + 2\beta(1 + \varepsilon)$ due to [Lemma 3.29](#) and ρ as in the proof of [Theorem 4.6](#). As before, we can reduce to the case $t = 0$.

Moreover, since J_B is only supported on the bulk, one can follow the proof of [Theorem 4.6](#) to argue that to bound $\mathbf{Cov}(\mu_{M_1+M_2,h})$, one only needs to bound covariances of Ising models on M_2 . Namely, it suffices to establish the following PSD upper bounds for arbitrary external field h' :

$$\begin{aligned} \|\mathbf{Cov}(\mu_{M_2,h'})_B\| &\leq O(1) \\ \|\mathbf{Cov}(\mu_{M_2,h'})_H\| &\leq O\left(\frac{\Delta}{D}\right). \end{aligned}$$

The main difficulty stems from the fact that $M_2 = J_R$ has support outside H ; it is now supported on the complete graph on H , as well as the edges between B and H .

To analyze $\mathbf{Cov}(\mu_{M_2,h})$, the high level strategy will be to use [Observation 3.9](#) and transfer over to the case of having interactions purely supported on H . To do so, we will prove an optimal Poincaré inequality for a Markov chain where the states are the connected components of H . Let us illustrate this argument informally; the subsequent lemmas will justify why we can execute these steps.

Recall that $\text{cc}(H)$ denotes the set of connected components of H , and let $C \in \text{cc}(H)$ be an arbitrary component. By a slight abuse of notation, let $M_C = (M_H)_C$ denote the restriction of M_H to C and $E_C = (E_H)_C$ denote the restriction for the error in the near-forest. For any $v \in \mathbb{R}^n$ supported on $S \subseteq [n]$, we have

$$\begin{aligned} \mathbf{Var}_{x \sim \mu_{M_2,h'}}(\langle v, x \rangle) &\leq O(1) \sum_{C \in \text{cc}(H)} \mathbf{E}_{x \sim \mu_{M_C+E_C,h(x-C)}} \mathbf{Var}(\langle v_C, x_C \rangle) && \text{(Poincaré)} \\ &\leq O(1) \sum_{C \in \text{cc}(H)} \mathbf{E}_{x \sim \mu_{M_C,h(x-C)}} \mathbf{Var}(\langle v_C, x_C \rangle) && \text{(Perturbation)} \\ &\leq O(1) \sum_{C \in \text{cc}(H)} \|v_C\|^2 \cdot \mathbf{E}_{x \sim \mu_{M_C,h(x-C)}} \left[\|\mathbf{Cov}(\mu_{M_C,h(x-C)})_S\| \right]. && (v \text{ supported on } S) \end{aligned}$$

By selecting $S = B$ and $S = H$, we can transfer over the covariance bounds computed for near-forest Ising models in [Lemma 4.5](#) to finish off the proof.

For the remainder of this section, we will justify the above steps and complete the proof of [Theorem 5.2](#).

Establishing an optimal Poincaré inequality. Let us first recall the definition of the Poincaré inequality on product domains.

Definition 5.18. Suppose that ν is a distribution on a product domain $\mathcal{P} = \mathcal{P}_1 \times \cdots \times \mathcal{P}_k$, and let $C_{\text{PI}}(\nu)$ be the Poincaré constant for Glauber dynamics on ν . Then for any $f : \mathcal{P} \rightarrow \mathbb{R}$, we have

$$\text{Var}_\nu(f) \leq \frac{1}{k \cdot C_{\text{PI}}(\nu)} \sum_{i \in [k]} \mathbf{E}_{y_{-i}} \text{Var}_{y_i | y_{-i}}(f).$$

We claim that the block Glauber dynamics studied in [Section 5.3](#) has an optimal Poincaré constant. In particular, for us $k = m + m'$ is the total number of connected components of H (including isolated vertices). Indeed, due to [Lemma 5.17](#) and [Corollary 5.11](#), we have

$$\begin{aligned} \|\mathcal{R}_\nu\| &\leq 8\alpha d^2 + \sqrt{8\alpha d^2} \\ &\leq \exp(-\Omega(\varepsilon^3 d)). \end{aligned} \tag{Lemma 5.15}$$

We now have the following lemma from [[Hay06](#)].

Lemma 5.19 ([\[Hay06\]](#)). *If the influence matrix of a distribution ν is bounded as $\|\mathcal{R}_\nu\| < 1 - \varepsilon$, the mixing time of Glauber dynamics is $O((n \log n) / \varepsilon)$.*

In conjunction with the previous equations, this implies an optimal mixing time bound for the block dynamics, and thus an $\Omega(\frac{1}{k})$ bound on $C_{\text{PI}}(\nu)$.

Perturbation. Next, we justify the transfer from $\mu_{M_C + E_C, h}$ to $\mu_{M_C, h}$, thereby removing the pesky additional weak interactions E_C . To accomplish this, we will use the simple perturbation bound [Lemma 3.8](#). Indeed, we already proved such a density bound for perturbed Ising models in [Lemma 5.13](#). Plugging in the explicit parameters for SBMs, as we did in the proof of [Lemma 5.17](#), we see that we can take $c = 1 + o_n(1)$ in [Lemma 3.8](#). In other words, we barely incur any loss transferring to M_C .

With the above lemmas in hand, we can finish off the proof of the theorem.

Proof of [Theorem 5.2](#). By [Observation 3.9](#), if $v \in \mathbb{R}^n$ is a unit norm vector supported on $S \subseteq [n]$, then there is a positive constant $c > 0$ such that

$$\begin{aligned} \text{Var}_{x \sim \mu_{M_R, h'}}(\langle v, x \rangle) &\leq c \sum_{C \in \text{cc}(H)} \mathbf{E}_{x_{-C}} \text{Var}_{x \sim \mu_{M_C + E_C, h}(x_{-C})}(\langle v_C, x_C \rangle) && \text{(Definition 5.18)} \\ &\leq c(1 + o_n(1)) \sum_{C \in \text{cc}(H)} \mathbf{E}_{x_{-C}} \text{Var}_{x \sim \mu_{M_C, h}(x_{-C})}(\langle v_C, x_C \rangle) && \text{(Lemma 3.8)} \\ &\leq c(1 + o_n(1)) \sum_{C \in \text{cc}(H)} \|v_C\|^2 \mathbf{E}_{x_{-C}} \left[\|\text{Cov}(\mu_{M_C, h}(x_{-C}))_S\| \right]. && (v \text{ supported on } S) \end{aligned}$$

Recall that $M_H = \frac{\beta}{\sqrt{d}}(\overline{A}_H \circ A_H)$. Since the centering interaction is $O(\frac{d}{n})$, it follows that the interactions are in $\left[-\frac{\beta}{\sqrt{d}}, \frac{\beta}{\sqrt{d}}\right] = \left[-\frac{\beta\sqrt{1+\varepsilon}}{\sqrt{d(1+\varepsilon)}}, \frac{\beta\sqrt{1+\varepsilon}}{\sqrt{d(1+\varepsilon)}}\right]$, so we can apply [Lemma 4.5](#) with $\gamma = \beta\sqrt{1+\varepsilon}$.

Taking $S = B$ and applying [Lemma 4.5](#) uniformly on each component C yields

$$\|\mathbf{Cov}(\mu_{M_R, h'})_B\| \leq c(1 + o(1)) \cdot \left(\frac{e^{2\beta/\sqrt{d}}}{(1 - \beta\sqrt{1+\varepsilon})^2} \right),$$

and for $S = H$ we obtain that

$$\|\mathbf{Cov}(\mu_{M_R, h'})_H\| \leq c(1 + o(1)) \cdot \left(\frac{e^{2\beta/\sqrt{d}}}{(1 - \beta\sqrt{1+\varepsilon})^2} \right) \cdot \frac{\Delta}{D}.$$

By following the same calculations as in [Theorem 4.1](#), the upshot is that for sufficiently small (constant) $\beta > 0$, we have for some constant C independent of Δ that

$$\mathbf{Cov}(\mu_{(1-t)J_B + J_R, h}) \preceq C (I_B + \Delta I_{H \setminus \partial H}),$$

as desired. ■

6 Near-forest decomposition for stochastic block models

In this section, we prove [Theorem 4.9](#), a statement about the existence of a near-forest decomposition for sparse random graphs.

Theorem 4.9. *Let $G \sim \text{SBM}(n, d, \lambda)$ for any $d \geq 1$. There exist universal constants C_1 and C_2 such that for any $\varepsilon \geq C_1 \cdot \left(\frac{C_2 + \log d}{d}\right)^{1/3}$, with high probability over G , we can partition the edge set $E(G)$ as $E(G_1) \sqcup E(G_2)$ for graphs G_1 and G_2 on $[n]$ where:*

1. *The maximum degree of G_1 is at most $(1 + \varepsilon)d$.*
2. *G_2 is a (Δ, D) -pseudorandom near-forest with respect to $S := V'(G_1) \cap V'(G_2)$ for $\Delta = o(\log n)$ and $D = (1 + \varepsilon)d$.*

Notation (ε, d, G) . Throughout this section, let ε and d be as in the statement of [Theorem 4.9](#), and let $G \sim \text{SBM}(n, d, \lambda)$.

We begin the proof by first defining the construction of G_1 and G_2 .

Definition 6.1 (Excision of a graph). For a graph G , and for each vertex $v \in V(G)$, define $N_\ell(v)$ to be the set of vertices at distance exactly ℓ from v and $B_\ell(v)$ to be the induced subgraph on G on those vertices within distance ℓ of v . We say that v is ℓ -heavy if $|B_\ell(v)| > (d(1 + \varepsilon))^\ell$, and conversely that v is ℓ -light if $|B_\ell(v)| \leq (d(1 + \varepsilon))^\ell$, and set

$$\ell(v) = \min\{\ell \geq 0 : v \text{ is } L\text{-light for all } L \geq \ell\}.$$

Note in particular that if $|B_L(v)| \leq (d(1 + \varepsilon))^L$ for all $L \geq 0$, then $\ell(v) = 0$ and $B_{\ell(v)}(v)$ has no edges. Finally, we define the *excision* of a graph G as

$$H := \bigcup_{v \in V} B_{\ell(v)}(v),$$

and refer to the graph after the removal of edges in H as the *bulk* of G .

The ingredients to prove [Theorem 4.9](#) are:

- [Observation 6.2](#), which says that the bulk of any graph has bounded degree, and the excision of any graph is (Δ, D) -pseudorandom for some Δ and $D = (1 + \varepsilon)d$.
- [Lemma 6.3](#), which says that the excision of a sparse random graph is a near-forest with high probability.
- [Lemma 6.4](#), which shows that $\Delta = o(\log n)$ with high probability.

Notation (H). We will use H to refer to the excision of G .

Observation 6.2. For any graph G , let G_1 and G_2 denote its bulk and excision respectively. Then, for some $\Delta > 0$ and $D := (1 + \varepsilon)d$:

1. The maximum degree of G_1 is at most D .
2. G_2 is (Δ, D) -pseudorandom with respect to $S := V'(G_1) \cap V'(G_2)$.

Lemma 6.3. *With probability $1 - o(1)$, H is a near-forest.*

The following statement establishes the claimed bound on Δ .

Lemma 6.4. *With probability $1 - o(1)$, for every vertex v in G and $\ell \geq 0$, $\frac{|B_\ell(v)|}{(d(1+\varepsilon))^\ell} = o(\log n)$.*

We are now ready to prove [Theorem 4.9](#).

Proof of [Theorem 4.9](#). Choose G_1 as the bulk of G , and G_2 as the excision of G . The statement then immediately follows from [Observation 6.2](#) and [Lemmas 6.3](#) and [6.4](#). ■

It now remains to prove the above statements. The statement of [Lemma 6.4](#) is a mild strengthening of [[BGGŠ24](#), Lemma 9 of arXiv version] and follows standard arguments involving branching processes, and hence, we defer its proof to [Appendix A](#).

The rest of this section is dedicated to proving [Lemma 6.3](#), which establishes that H is a near forest with high probability.

6.1 Excisions of sparse random graphs are near-forests

The key technical ingredient in the proof of [Lemma 6.3](#) is the following.

Lemma 6.5. *With probability $1 - o(1)$, for any connected component C of H , $\text{diam}(C) \leq O\left(\frac{\log n}{\varepsilon^3 d}\right)$.*

As foreshadowed in the technical overview, we will also need the following structural result about small sets in sparse random graphs.

Lemma 6.6 ([\[BLM15, Lemma 30\]](#), also see [\[BGGŠ24, Lemma 31\]](#)). *For $G \sim \text{SBM}(n, d, \lambda)$, with high probability, there is no connected set S of vertices with $|S| < \frac{1}{2} \cdot \frac{\log n}{\log d}$ that has tree-excess larger than 1.*

Assuming the above, let us present a proof of [Lemma 6.3](#).

Proof of [Lemma 6.3](#). Condition on the event in [Lemma 6.5](#), so we have that every component C of H has diameter at most $\log n/d$. Suppose instead that C had tree-excess at least 2, and let C_1, C_2 be minimal cycles in C (in that they do not have any chords). Consider the induced subgraph on the set S of vertices in C comprised of C_1, C_2 , and a shortest path (in C) between the two. Note in particular that S is connected, and $|S| \leq 5 \text{diam}(C)$. Since $\varepsilon = \Omega_d(\log d/d)^{1/3}$, we may apply [Lemma 6.6](#) to conclude that the tree-excess of the induced subgraph on S is at most 1, a contradiction. ■

We now turn our attention to proving [Lemma 6.5](#), which bounds the diameter of each connected component of H . A key auxiliary graph in our proof is the *cluster graph*.

Definition 6.7. Given a graph G on $[n]$, the associated *cluster graph* $\text{Cluster}(G)$ has vertex set $[n]$, with an edge between u and v iff $\ell(u), \ell(v) > 0$ and $B_{\ell(u)}(u) \cap B_{\ell(v)}(v) \neq \emptyset$. Furthermore, each vertex u in this graph has associated weight $\text{wt}(u) = (1 + \varepsilon)^{\ell(u)}$, and each edge uv has weight $\text{wt}(u, v) = \ell(u) + \ell(v)$.

Observe that [Lemma 6.5](#) follows immediately from the following two statements:

1. [Lemma 6.8](#), which shows that if a connected component of H has large diameter, then it causes a high-weight induced path in $\text{Cluster}(G)$.
2. [Lemma 6.9](#), which establishes that there are no high-weight induced paths in $\text{Cluster}(G)$.

Lemma 6.8. *Suppose the diameter of a connected component of H exceeds τ . Then there are vertices $u_1, \dots, u_k \in [n]$ such that the induced subgraph in $\text{Cluster}(G)$ is the path $u_1 \dots u_k$, and*

$$\sum_{i \in [k]} \ell(u_i) \geq \frac{\tau}{2}.$$

Lemma 6.9. *Define η as $\min\left\{\frac{\varepsilon^2}{54}, \frac{1}{18}\right\}$. With probability $1 - o(1)$, for any $U = \{u_1, \dots, u_k\}$ such that the induced subgraph on U in $\text{Cluster}(G)$ is the path $u_1 \dots u_k$, we have:*

$$\sum_{i \in [k]} \ell(u_i) < \frac{32 \log n}{\eta \varepsilon d}.$$

The proof of [Lemma 6.8](#) is fairly short and self-contained, and we present it below.

Proof of [Lemma 6.8](#). There must be vertices a and b in a single connected component of H such that $\text{dist}_H(a, b) > \tau$. Let u and v be vertices such that $B_{\ell(a)}(a)$ and $B_{\ell(b)}(b)$ contain u and v respectively.

Let $u_1 u_2 \dots u_k$ be the weighted shortest path between u and v in $\text{Cluster}(\mathbf{G})$ where $u_1 = u$ and $u_k = v$. The path between a and b has length bounded by $2 \sum_{i \in [k]} \ell_i$, and hence:

$$\sum_{i \in [k]} \ell_i \geq \frac{\tau}{2}.$$

The induced subgraph on $\{u_1, \dots, u_k\}$ certainly contains the claimed path by construction. It remains to show that there are no other edges besides the ones in the path. If there is an edge between u_i and u_j for $j \neq i \pm 1$ for $i < j$, then the path $u_1 \dots u_i u_j u_{j+1} \dots u_k$ is a shorter path than the original one, which leads to a contradiction. Indeed, since the edge weights are positive and $\text{wt}(u, v) = \ell(u) + \ell(v)$, any path between u_i and u_j , in particular $u_i u_{i+1} \dots u_{j-1} u_j$, has weight at least $\ell(u_i) + \ell(u_j)$, with equality if and only if the path is just the edge $u_i u_j$. ■

We now begin the proof of [Lemma 6.9](#) by setting up some terminology.

Definition 6.10 (Violating set and minimal violating set). We say a subset of vertices U is a *violating set* if the induced subgraph on U in $\text{Cluster}(\mathbf{G})$ is a path of weight exceeding $\frac{32 \log n}{\eta \epsilon d}$. We say U is a *minimal violating set* if it is a violating set, and no subset of U is a violating set.

Definition 6.11 (Signature). The *signature* of a minimum violating set is the tuple $(k, \vec{u}, \vec{\ell})$ where k is the number of vertices in the minimum violating set, $\vec{u} = (u_1, \dots, u_k)$ is the vertices in the minimum violating set in the order they appear in the path, and $\vec{\ell} = (\ell(u_1), \dots, \ell(u_k))$.

Notation (\mathcal{S}_κ). For any positive integer κ , define \mathcal{S}_κ as all signatures $(k, \vec{u}, \vec{\ell})$ for which $k \geq 1$ and $\vec{\ell}$ is in $[\kappa]^k$.

To prove [Lemma 6.9](#), it suffices to prove that there are no violating sets in $\text{Cluster}(\mathbf{G})$. Our goal is to prove that with high probability, there are no violating sets in $\text{Cluster}(\mathbf{G})$. We observe that it suffices to prove there are no minimal violating sets ([Observation 6.12](#)), and then bound the probability of containing a minimal violating set in [Observation 6.13](#).

Observation 6.12. If $U \subseteq [n]$ is a violating set in $\text{Cluster}(\mathbf{G})$, then it contains a minimal violating set $U' \subseteq U$.

Observation 6.13. For any positive integer κ , the probability that $\text{Cluster}(\mathbf{G})$ contains a minimal violating set is bounded by

$$\sum_{(k, \vec{u}, \vec{\ell}) \in \mathcal{S}_\kappa} \Pr \left[\exists \text{MVS with signature } (k, \vec{u}, \vec{\ell}) \text{ in } \text{Cluster}(\mathbf{G}) \right] + \Pr[\exists v \in [n] \text{ s.t. } \ell(v) > \kappa],$$

where MVS is short for “minimal violating set”.

The following, proved in [Appendix A](#), gives a choice of κ for which the second term of the above is negligible.

Lemma 6.14. For $\kappa = \frac{4}{\varepsilon} \log \frac{\log n}{\eta d}$, we have:

$$\Pr[\exists v \in [n] \text{ s.t. } \ell(v) > \kappa] = o(1).$$

The following gives us a handle on the first term of the bound in [Observation 6.13](#).

Lemma 6.15. For $\kappa = \frac{4}{\varepsilon} \log \frac{\log n}{\eta d}$ and for any $(k, \vec{u}, \vec{\ell}) \in \mathcal{S}_k$ where $\vec{\ell} = (\ell_1, \dots, \ell_k)$, we have:

$$\begin{aligned} & \Pr[\exists \text{MVS with signature } (k, \vec{u}, \vec{\ell}) \text{ in Cluster}(G)] \\ & \leq \frac{1}{n^{k-1}} \prod_{i \in [k]} 2\ell_i^2 \cdot (d(1+\varepsilon))^{2\ell_i} \cdot \exp\left(-\frac{\eta d}{2} \left(1 + \frac{\varepsilon}{4}\right)^{\ell_i}\right). \end{aligned}$$

Remark 6.16. The way to interpret the above is to think of the probability bound as decaying at a *doubly* exponential rate in each ℓ_i , and for the events $\{u_i \text{ is } \ell_i\text{-heavy}\}$ as being approximately independent. See [Lemma A.1](#) and its proof to understand the doubly exponential decay rate.

We first show how to complete the proof of [Lemma 6.9](#), and then later prove [Lemma 6.15](#).

Proof of Lemma 6.9. By [Observation 6.12](#) it suffices to prove that there is no minimal violating set in $\text{Cluster}(G)$. The definition of a signature $(k, \vec{u}, \vec{\ell})$ of a minimal violating set implies that $\sum_i \ell_i \geq \frac{32 \log n}{\eta \varepsilon d}$. Hence, by [Observation 6.13](#) and [Lemmas 6.14](#) and [6.15](#), we get that the probability of containing a violating set is at most:

$$n \sum_{k \geq 1} \sum_{\substack{\ell_1, \dots, \ell_k \\ \sum \ell_i \geq \frac{32 \log n}{\eta \varepsilon d}}} \prod_{i \in [k]} 2\ell_i^2 \cdot (d(1+\varepsilon))^{2\ell_i} \cdot \exp\left(-\frac{\eta d}{2} \left(1 + \frac{\varepsilon}{4}\right)^{\ell_i}\right). \quad (11)$$

Consider one of the terms in the product, which is

$$\exp\left(\log(2\ell_i^2) + 2\ell_i \log(d(1+\varepsilon)) - \frac{\eta d}{2} \left(1 + \frac{\varepsilon}{4}\right)^{\ell_i}\right) \leq \exp\left(\ell_i(2 \log(d(1+\varepsilon)) + 2) - \frac{\eta d}{2} \left(1 + \frac{\varepsilon}{4}\right)^{\ell_i}\right).$$

We claim that for

$$\varepsilon \geq 1500 \left(\frac{100 + \log d}{d}\right)^{1/3}, \quad (12)$$

the above is at most $\exp\left(-\ell_i \left(1 + \frac{\eta \varepsilon d}{16}\right)\right)$. Indeed, since $\left(1 + \frac{\varepsilon}{4}\right)^{\ell_i} \geq \frac{\varepsilon}{4} \cdot \ell_i$, the above quantity is at most

$$\exp\left(-\ell_i \left(\frac{\eta \varepsilon d}{8} - 2 - 2 \log(d(1+\varepsilon))\right)\right).$$

Therefore, it suffices for ε to satisfy

$$\frac{\eta \varepsilon d}{16} \geq 2 \log(d(1+\varepsilon)) + 3.$$

It is straightforward to verify that ε as in (12) satisfies the above. Given this, (11) is at most

$$\begin{aligned}
n \sum_{k \geq 1} \sum_{\substack{\ell_1, \dots, \ell_k \\ \sum \ell_i \geq \frac{32 \log n}{\eta \varepsilon d}}} \exp \left(- \left(\frac{\eta \varepsilon d}{16} + 1 \right) \sum_{i \in [k]} \ell_i \right) &\leq \frac{1}{n} \sum_{k \geq 1} \sum_{\substack{\ell_1, \dots, \ell_k \\ \sum \ell_i \geq \frac{32 \log n}{\eta \varepsilon d}}} \exp \left(- \sum_{i \in [k]} \ell_i \right) \\
&\leq \frac{1}{n} \sum_{k \geq 1} \sum_{\ell_1, \dots, \ell_k \geq 1} \exp \left(- \sum_{i \in [k]} \ell_i \right) \\
&\leq \frac{1}{n} \sum_{k \geq 1} \exp(-\Omega(k)) = o(1),
\end{aligned}$$

as desired. ■

We now turn our attention to proving [Lemma 6.15](#). The following two statements ([Lemmas 6.17](#) and [6.19](#)) will be used in the proof. We defer their proofs to [Appendix A](#).

Lemma 6.17. *Let $u_1, u_2, \dots, u_r \in [n]$ and $\ell_1, \ell_2, \dots, \ell_r > 0$. Then,*

$$\Pr \left[\ell(u_i) = \ell_i \ \forall i \in [r] \text{ and } B_{\ell_i}(u_i) \cap B_{\ell_j}(u_j) = \emptyset \ \forall i \neq j \in [r] \right] \leq \prod_{i \in [r]} \exp \left(-\eta d \left(1 + \frac{\varepsilon}{4} \right)^{\ell_i} \right).$$

Remark 6.18. In the above bound, the reader should treat the events $\{\ell(u_i) = \ell_i\}$ as roughly independent. Informally, the approximate independence can be justified by the fact that the union of the balls covers only a small fraction of the graph, and hence for a vertex u_r disjoint from $B_{\ell(u_1)}(u_1) \cup \dots \cup B_{\ell(u_{r-1})}(u_{r-1})$, the probability that $B_{\ell(u_r)}(u_r)$ intersects the rest is very small — it is roughly the probability that two small subsets of $[n]$ intersect. Furthermore, when the ball around u_r is disjoint from the remaining balls, its size is not influenced by the other balls, resulting in approximate independence.

Lemma 6.19. *Let S, T_1, T be disjoint subsets of vertices, and say $T = T_2 \sqcup T'$. Let \mathcal{E} be an (arbitrary) conditioning of the edges adjacent to vertices in S, T_1, T such that for any $A = S, T_1, T$, we can write $A = A_{\text{in}} \sqcup A_{\text{out}}$ such that the following holds.*

- $|A| = o(n)$.
- Every edge incident on vertices in A_{in} is fixed in a way that there are no edges between A_{out} and A^c .
- Further, there is no conditioning done on the edges between A_{out} and $(S_{\text{in}} \cup (T_1)_{\text{in}} \cup T_{\text{in}})^c$.

Let $u \in [n]$ and $\ell > 0$. Then, for n sufficiently large,

$$\Pr [u \text{ is } \ell\text{-light and } B_\ell(u) \cap T_i \neq \emptyset \text{ for } i \in \{1, 2\} \mid \mathcal{E}] \leq \frac{2\ell^2 \cdot (d(1 + \varepsilon))^{2\ell} \cdot |T_1| \cdot |T_2|}{n^2}.$$

Remark 6.20. Put more intuitively, [Lemma 6.19](#) is a generalization of the following simple observation: for any small set of vertices of size C , then the probability that a random vertex is contained in said set scales like C/n . The technical assumptions on S, T_1 , and T are present to ensure that the lemma applies throughout the proof of [Lemma 6.9](#); see [Figure 4](#) for a depiction of these sets.

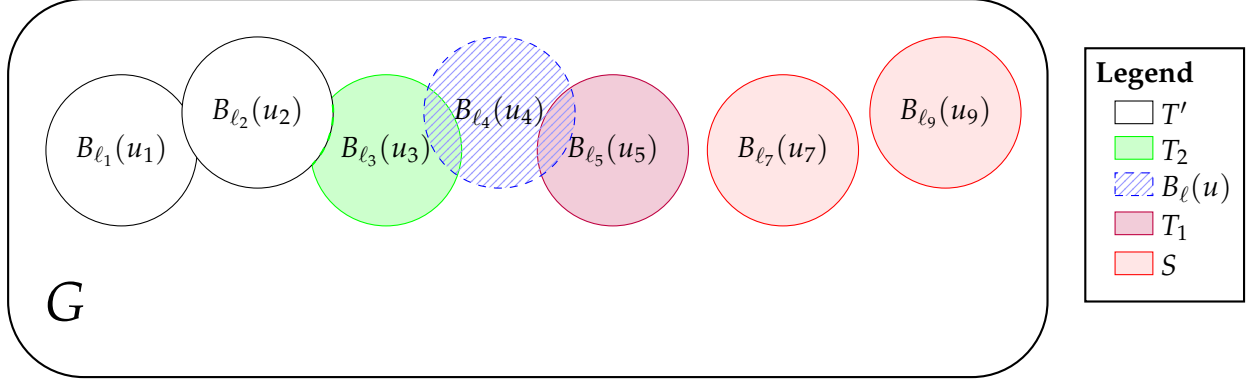


Figure 4: A depiction of the sets T_1, T_2, T', S , and $B_{\ell}(u)$. The event of interest in [Lemma 6.19](#) is when the random set of interest $B_{\ell}(u)$ (shaded blue) intersects the two target sets T_1 (solid purple) and T_2 (solid green). The other two sets of vertices S and T' cover the additional edges that need to be conditioned on to apply the lemma inductively to bound the length of the path $u_1 \dots u_k$ in the cluster graph. In our setting, all the sets involved are unions of balls centered at the vertices u_i .

We now prove [Lemma 6.15](#).

Proof of Lemma 6.15. Assume, without loss of generality, that

$$\sum_{i \text{ odd}} (1 + \varepsilon)^{\ell_i} \geq \sum_{i \text{ even}} (1 + \varepsilon)^{\ell_i}.$$

Thus, for the independent set $\{u_i : i \text{ odd}\}$, we have

$$\sum_{i \text{ odd}} (1 + \varepsilon)^{\ell_i} \geq \frac{1}{2} \sum_i (1 + \varepsilon)^{\ell_i}.$$

Let $\mathcal{E}_{\text{independent}}$ be the event that $\ell(u_i) = \ell_i$ for all odd i and $B_{\ell_i}(u_i) \cap B_{\ell_j}(u_j) = \emptyset$ for distinct odd i, j . Let $\mathcal{E}_{\text{cross}}$ be the event that for all even i and $j = i \pm 1$, $B_{\ell_i}(u_i) \cap B_{\ell_j}(u_j) \neq \emptyset$. Then,

$$\begin{aligned} & \Pr \left[\exists \text{MVS with signature } (k, \vec{u}, \vec{\ell}) \text{ in Cluster}(\mathbf{G}) \right] \\ & \leq \Pr[\mathcal{E}_{\text{independent}}, \mathcal{E}_{\text{cross}}] \end{aligned} \quad (13)$$

$$\begin{aligned} & = \Pr[\mathcal{E}_{\text{independent}}] \cdot \Pr[\mathcal{E}_{\text{cross}} \mid \mathcal{E}_{\text{independent}}] \\ & \leq \prod_{i \text{ odd}} \exp(-\eta d (1 + \varepsilon)^{\ell_i}) \cdot \prod_{i \text{ even}} 2\ell_i^2 \cdot \frac{(d(1 + \varepsilon))^{2\ell_i + \ell_{i-1} + \ell_{i+1}}}{n^2} \end{aligned} \quad (14)$$

$$\leq \frac{1}{n^{k-1}} \prod_{i \in [k]} 2\ell_i^2 \cdot (d(1 + \varepsilon))^{2\ell_i} \cdot \exp\left(-\frac{\eta d}{2} \left(1 + \frac{\varepsilon}{4}\right)^{\ell_i}\right). \quad (15)$$

To explain the above chain of inequalities, in (13) we dropped the condition that $\ell(u_i) = \ell_i$ for even i . In (14), we applied [Lemma 6.17](#) and [Lemma 6.19](#) inductively with the setting

$$T_1 = B_{\ell_{i+1}}(u_{i+1}), S = \bigcup_{\substack{j \text{ odd} \\ j > i+1}} B_{\ell_j}(u_j), T' = \bigcup_{j < i-1} B_{\ell_j}(u_j), T_2 = B_{\ell_{i-1}}(u_{i-1}) \setminus T'.$$

When applying [Lemma 6.19](#), we use the bound of κ on ℓ_i to conclude that the size of the union of all these balls is $o(n)$. See [Figure 4](#) for an illustration. In [\(15\)](#), we used the fact that the odd vertices have at least half of the total vertex weight. \blacksquare

6.2 Bounding the sizes of components

In this section, we prove high probability bounds on the sizes of connected components in H , which are of interest when studying the Ising model on centered interaction matrices.

[Lemma 5.16](#), which states that the maximum size of a connected component is $n^{o_d(1)}$, is a consequence of [Lemma 6.5](#).

Lemma 5.16. *With probability $1 - o(1)$, the maximum size of a connected component is at most $n \frac{c \log(d(1+\varepsilon))}{\varepsilon^3 d}$ for some absolute constant $c > 0$.*

We believe that our analysis here is loose, and that the size of the largest connected component is in fact poly $\log(n)$ with high probability.

Proof. By [Lemma 6.5](#), we know that for any connected component C , there is an absolute positive constant c such that $\text{diam}(C) \leq \frac{c \log n}{\varepsilon^3 d}$. Note that any vertex $u \in C$ is $\text{diam}(C)$ -light. Furthermore, we have $B_{\text{diam}(C)}(u) \supseteq C$, so

$$|C| \leq (d(1 + \varepsilon))^{\text{diam}(C)} \leq \exp\left(\log(d(1 + \varepsilon)) \cdot \frac{c \log n}{\varepsilon^3 d}\right) = n \frac{c \log(d(1+\varepsilon))}{\varepsilon^3 d},$$

as desired. \blacksquare

Using similar ingredients, we are also able to establish the following quantitative bound on the sum of squared component sizes, from which [Lemma 5.15](#) follows.

Lemma 6.21. *Write $H = \sqcup_{C \in \text{cc}(H)} C$, where $\text{cc}(H)$ denotes the connected components of H . There are absolute constants $c_1, c_2 > 0$ such that for $\varepsilon \geq c_1(\log d/d)^{1/3}$, with probability $1 - o(1)$,*

$$\sum_{\substack{C \in \text{cc}(H) \\ |C| > 1}} |C|^2 \leq \exp(-c_2 \varepsilon^3 d) n.$$

We now turn our attention to proving [Lemma 6.21](#), which gives a high probability bounds on the sizes of component sizes in H . The proof essentially amounts to bounding the first and second moments. For the first moment bound, many of the key ingredients are already contained in the proof of [Lemma 6.5](#); for the variance bound, we use the Efron-Stein inequality.

Proof of [Lemma 6.21](#). For succinctness, define the random variable

$$X := \sum_{C \in \text{cc}(H)} |C|^2 \cdot \mathbf{1}[|C| > 1].$$

For distinct $u, v \in [n]$, define the random variable $X_{uv} = \mathbf{1}[u, v \text{ are in the same component}]$. Notice that

$$X = \sum_{u \neq v} X_{uv}.$$

Hence for the first moment it suffices to upper bound $\mathbf{E}[X_{uv}] = \Pr[u, v \text{ are in the same component}]$. In fact, bounding $\mathbf{E}[X_{uv}]$ uses tools similar to what we used in the proof of [Lemma 6.5](#).

Observe that for distinct u and v , they are in the same connected component only if there are u' and v' (which need not be distinct) which satisfy the following:

- $\ell_{u'}, \ell_{v'} > 0$, so that u' and v' belong to nontrivial connected components.
- $u \in B_{\ell_{u'}}(u')$ and $v \in B_{\ell_{v'}}(v')$.
- There is a path between u' and v' in $\text{Cluster}(\mathbf{G})$.

Note that if $u' = v'$, we allow for self loops, as long as $\ell_{u'} > 0$; this is consistent with [Definition 6.7](#). Hence, we have

$$\sum_{u \neq v} X_{uv} \leq \sum_{u', v' \in [n]} |B_{\ell_{u'}}(u')| |B_{\ell_{v'}}(v')| \mathbf{1}[u', v' \text{ connected in Cluster}(\mathbf{G})].$$

Fix a path length k , positive integers ℓ_i for $i \in [k]$, and a path $w_1 \cdots w_k$ with $w_1 = u'$ and $w_k = v'$. Consider the event \mathcal{E} whose probability we bounded in [Lemma 6.9](#):

$$\mathcal{E} := \{\ell_i = \ell(w_i) \text{ for } i \in [k] \text{ and } B_{\ell_i}(w_i) \cap B_{\ell_j}(w_j) \neq \emptyset \text{ iff } j = i \pm 1 \text{ for } i \in [k]\}$$

For the term at u', v' in the sum to be nonzero, the above event \mathcal{E} must occur for some choice of k , positive ℓ_i , and path from u' to v' . On \mathcal{E} , we can upper bound

$$|B_{\ell_{u'}}(u')| \cdot |B_{\ell_{v'}}(v')| \cdot \mathbf{1}[u', v' \text{ connected in Cluster}(\mathbf{G})] \leq (d(1 + \varepsilon))^{\ell_1 + \ell_k},$$

and a careful inspection of the proof of [Lemma 6.9](#) establishes that

$$(d(1 + \varepsilon))^{\ell_1 + \ell_k} \Pr[\mathcal{E}] \leq \prod_{i \in [k]} 2\ell_i^2 \cdot (d(1 + \varepsilon))^{2\ell_i} \cdot \exp\left(-\frac{\eta d}{2} \left(1 + \frac{\varepsilon}{4}\right)^{\ell_i}\right).$$

Indeed, in [\(14\)](#), only one factor of $(d(1 + \varepsilon))^{\ell_i}$ is picked up for $i \in \{1, k\}$, as they have only one neighbor in the chain.

Now choose ε such that

$$\frac{\eta \varepsilon d}{16} \geq 2 \log(d(1 + \varepsilon)) + 2,$$

which is satisfied for $\varepsilon \geq 1500 \left(\frac{100 + \log d}{d}\right)^{1/3}$. Summing up now over all k, ℓ_i , and paths from u' to v' , we obtain

$$\mathbf{E} \left[\sum_{u', v' \in [n]} |B_{\ell_{u'}}(u')| \cdot |B_{\ell_{v'}}(v')| \cdot \mathbf{1}[u', v' \text{ connected in Cluster}(\mathbf{G})] \right]$$

$$\begin{aligned}
&\leq n \sum_{k \geq 1} \sum_{\ell_1, \dots, \ell_k \geq 1} \prod_{i \in [k]} 2\ell_i^2 \cdot (d(1+\varepsilon))^{2\ell_i} \cdot \exp\left(-\frac{\eta d}{2}(1+\frac{\varepsilon}{4})^{\ell_i}\right) \\
&\leq n \sum_{k \geq 1} \sum_{\ell_1, \dots, \ell_k \geq 1} \exp\left(-\frac{\eta \varepsilon d}{16} \sum_{i \in [k]} \ell_i\right) \\
&\leq n \sum_{k \geq 1} \exp\left(-\frac{\eta \varepsilon k d}{16}\right) \prod_{i \in [k]} \frac{\exp\left(-\frac{\eta \varepsilon d}{16}\right)}{1 - \exp\left(-\frac{\eta \varepsilon d}{16}\right)} \\
&\leq n \sum_{k \geq 1} \exp\left(-\frac{\eta \varepsilon k d}{16}\right) \\
&\leq 2 \exp\left(-\frac{\eta \varepsilon d}{16}\right) n,
\end{aligned}$$

where in the second-to-last line we have used the fact that $\frac{\eta \varepsilon d}{16} \geq 2$ and $\frac{\exp(-2)}{1 - \exp(-2)} \leq 1$, and in the last line we have used $\frac{1}{1 - \exp(-2)} \leq 2$. Hence, $\mathbf{E}[X] \leq 2 \exp\left(-\frac{\eta \varepsilon d}{16}\right) n$.

We now show that $\mathbf{Var}(X) = o(\mathbf{E}[X]^2)$, which implies the desired result by applying Chebyshev. To establish this, we appeal to the Efron-Stein inequality. Consider rerandomizing an edge e in G . With probability at most $1 - 2d/n$, the rerandomized edge e agrees with the original edge, so the value of X does not change. If it disagrees, then X changes only if e connects two connected components or if e breaks a connected component into two smaller connected components. Either way, we can upper bound the magnitude of the difference by $\max_{C \in \text{cc}(H)} 2|C|^2$.

By [Lemma 5.16](#), $\max_{C \in \text{cc}(H)} |C|^2 \leq n \frac{c \log d}{\varepsilon^3 d}$ for some absolute constant $c > 0$, so

$$\begin{aligned}
\mathbf{Var}(X) &\leq \frac{4d}{n} \cdot n \frac{c \log d}{\varepsilon^3 d} \cdot n^2 \\
&\leq 4 \exp\left(\log d + \frac{c \log d}{\varepsilon^3 d} \log n\right) n.
\end{aligned}$$

To conclude, it suffices to verify that for our choice of parameters

$$\log d + \frac{c \log d}{\varepsilon^3 d} \log n < -\varepsilon^3 d + \frac{1}{2} \log n.$$

Indeed, for $\varepsilon \geq \left(\frac{K \log d}{d}\right)^{1/3}$ and any $d = O(\log n)$, the left-hand-side of the above inequality is $O(\log \log n) + \frac{c}{K} \log n$, whereas the right-hand-side is $\frac{1}{2} \log n - O(\log \log n)$. Hence, the above inequality is satisfied for K a sufficiently large constant, which finishes the proof. \blacksquare

Acknowledgments

S.M. would like to thank Frederic Koehler for helpful conversations about Ising models.

References

- [AASV21] Yeganeh Alimohammadi, Nima Anari, Kirankumar Shiragur, and Thuy-Duong Vuong. Fractionally log-concave and sector-stable polynomials: counting planar matchings and more. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 433–446, 2021. [39](#)
- [Abb18] Emmanuel Abbe. Community Detection and Stochastic Block Models: Recent Developments. *Journal of Machine Learning Research*, 18(177):1–86, 2018. [3](#)
- [AHL⁺23] Nima Anari, Yizhi Huang, Tianyu Liu, Thuy-Duong Vuong, Brian Xu, and Katherine Yu. Parallel discrete sampling via continuous walks. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pages 103–116, 2023. [4](#)
- [AJK⁺21a] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Entropic independence I: Modified log-Sobolev inequalities for fractionally log-concave distributions and high-temperature Ising models. *arXiv preprint arXiv:2106.04105*, 2021. [5](#), [39](#)
- [AJK⁺21b] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Entropic Independence II: Optimal Sampling and Concentration via Restricted Modified Log-Sobolev Inequalities. *arXiv preprint arXiv:2111.03247*, 2021. [1](#)
- [AJK⁺24] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Universality of Spectral Independence with Applications to Fast Mixing in Spin Glasses. In *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 5029–5056. SIAM, 2024. [1](#), [39](#)
- [AKV24] Nima Anari, Frederic Koehler, and Thuy-Duong Vuong. Trickle-Down in Localization Schemes and Applications. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1094–1105, 2024. [2](#)
- [AS15] Emmanuel Abbe and Colin Sandon. Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 670–688. IEEE, 2015. [3](#)
- [BB19] Roland Bauerschmidt and Thierry Bodineau. A very simple proof of the LSI for high temperature spin systems. *Journal of Functional Analysis*, 276(8):2582–2588, 2019. [1](#), [4](#)
- [BBD23] Roland Bauerschmidt, Thierry Bodineau, and Benoit Dagallier. Kawasaki dynamics beyond the uniqueness threshold. *arXiv preprint arXiv:2310.04609*, 2023. [1](#), [4](#)
- [BCC⁺22] Antonio Blanca, Pietro Caputo, Zongchen Chen, Daniel Parisi, Daniel Štefankovič, and Eric Vigoda. On Mixing of Markov Chains: Coupling, Spectral Independence, and Entropy Factorization. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 3670–3692. SIAM, 2022. [39](#)

- [BD97] Russ Bubley and Martin Dyer. Path coupling: A technique for proving rapid mixing in Markov chains. In *Proceedings 38th Annual Symposium on Foundations of Computer Science*, pages 223–231. IEEE, 1997. [1](#), [39](#)
- [BGGŠ24] Ivona Bezáková, Andreas Galanis, Leslie Ann Goldberg, and Daniel Štefankovič. Fast Sampling via Spectral Independence Beyond Bounded-degree Graphs. *ACM Trans. Algorithms*, 20(1), Jan 2024. [47](#), [48](#)
- [BLM15] Charles Bordenave, Marc Lelarge, and Laurent Massoulié. Non-backtracking spectrum of random graphs: community detection and non-regular Ramanujan graphs. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 1347–1357. IEEE, 2015. [3](#), [12](#), [21](#), [48](#)
- [BT06] Sergey G. Bobkov and Prasad Tetali. Modified logarithmic sobolev inequalities in discrete settings. *Journal of Theoretical Probability*, 19(2):289–336, Jun 2006. [5](#), [16](#)
- [CE22] Yuansi Chen and Ronen Eldan. Localization Schemes: A Framework for Proving Mixing Bounds for Markov Chains. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 110–122, Los Alamitos, CA, USA, Nov 2022. IEEE Computer Society. [1](#), [4](#), [5](#), [7](#), [19](#)
- [CLV21] Zongchen Chen, Kuikui Liu, and Eric Vigoda. Optimal mixing of Glauber dynamics: entropy factorization via high-dimensional expansion. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2021*, page 1537–1550. Association for Computing Machinery, 2021. [1](#)
- [DKMZ11] Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E*, 84:066106, Dec 2011. [3](#)
- [Dob68] PL Dobruschin. The description of a random field by means of conditional probabilities and conditions of its regularity. *Theory of Probability & Its Applications*, 13(2):197–224, 1968. [1](#), [39](#)
- [DSS23] Jian Ding, Jian Song, and Rongfeng Sun. A new correlation inequality for Ising models with external fields. *Probability Theory and Related Fields*, 186(1):477–492, 2023. [28](#)
- [EAM22] Ahmed El Alaoui and Andrea Montanari. An information-theoretic view of stochastic localization. *IEEE Transactions on Information Theory*, 68(11):7423–7426, 2022. [7](#)
- [EAMS22] 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. [4](#)
- [EAMS23] Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Sampling from Mean-Field Gibbs Measures via Diffusion Processes. *arXiv preprint arXiv:2310.08912*, 2023. [4](#)

- [EF23] 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 Kousha Etessami, Uriel Feige, and Gabriele Puppis, editors, *50th International Colloquium on Automata, Languages, and Programming (ICALP 2023)*, volume 261 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 54:1–54:17, Dagstuhl, Germany, 2023. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. [4](#)
- [EKPS00] William Evans, Claire Kenyon, Yuval Peres, and Leonard J. Schulman. Broadcasting on trees and the Ising model. *Ann. Appl. Probab.*, 10(2):410–433, 2000. [1](#)
- [EKZ22] Ronen Eldan, Frederic Koehler, and Ofer Zeitouni. A spectral condition for spectral gap: fast mixing in high-temperature Ising models. *Probability theory and related fields*, 182(3):1035–1051, 2022. [1](#), [4](#), [5](#), [7](#)
- [Eld13] Ronen Eldan. Thin shell implies spectral gap up to polylog via a stochastic localization scheme. *Geometric and Functional Analysis*, 23(2):532–569, March 2013. [4](#), [19](#)
- [EZ24] Charilaos Efthymiou and Kostas Zampetakis. On sampling diluted Spin Glasses using Glauber dynamics. *arXiv preprint arXiv:2403.08921*, 2024. [4](#)
- [FM17] Zhou Fan and Andrea Montanari. How well do local algorithms solve semidefinite programs? In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 604–614, 2017. [21](#)
- [FV17] Sacha Friedli and Yvan Velenik. *Statistical Mechanics of Lattice Systems. A Concrete Mathematical Introduction*. 2017. [1](#)
- [Goe04] Sharad Goel. Modified logarithmic Sobolev inequalities for some models of random walk. *Stochastic processes and their applications*, 114(1):51–79, 2004. [16](#)
- [Gri06] Geoffrey Grimmett. *The random-cluster model*, volume 333. Springer, 2006. [28](#)
- [GT04] Francesco Guerra and Fabio Lucio Toninelli. The high temperature region of the Viana–Bray diluted spin glass model. *Journal of statistical physics*, 115:531–555, 2004. [2](#)
- [Hay06] Thomas P Hayes. A simple condition implying rapid mixing of single-site dynamics on spin systems. In *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 39–46. IEEE, 2006. [38](#), [45](#)
- [HMP24] 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. [4](#)
- [HS86] Richard Holley and Daniel W Stroock. Logarithmic Sobolev inequalities and stochastic Ising models. 1986. [16](#)

- [JS93] Mark Jerrum and Alistair Sinclair. Polynomial-time approximation algorithms for the Ising model. *SIAM Journal on computing*, 22(5):1087–1116, 1993. [1](#)
- [KLR22] Frederic Koehler, Holden Lee, and Andrej Risteski. Sampling approximately low-rank Ising models: MCMC meets variational methods. In *Conference on Learning Theory*, pages 4945–4988. PMLR, 2022. [1](#)
- [Kun23] Dmitriy Kunisky. Optimality of Glauber dynamics for general-purpose Ising model sampling and free energy approximation. *arXiv preprint arXiv:2307.12581*, 2023. [2](#)
- [Led01] Michel Ledoux. Logarithmic Sobolev inequalities for unbounded spin systems revisited. *Séminaire de Probabilités XXXV*, pages 167–194, 2001. [16](#)
- [Liu21] Kuikui Liu. From Coupling to Spectral Independence and Blackbox Comparison with the Down-Up Walk. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2021)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2021. [39](#)
- [LMR22] Siqi Liu, Sidhanth Mohanty, and Prasad Raghavendra. On statistical inference when fixed points of belief propagation are unstable. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 395–405. IEEE, 2022. [21](#)
- [LMR⁺24] Kuikui Liu, Sidhanth Mohanty, Prasad Raghavendra, Amit Rajaraman, and David X. Wu. Locally Stationary Distributions: A Framework for Analyzing Slow-Mixing Markov Chains, 2024. [0](#), [3](#)
- [Mas14] Laurent Massoulié. Community detection thresholds and the weak Ramanujan property. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 694–703, 2014. [3](#)
- [MM09] Marc Mézard and Andrea Montanari. *Information, Physics, and Computation*. Oxford University Press, 01 2009. [1](#)
- [MNS14] Elchanan Mossel, Joe Neeman, and Allan Sly. Belief propagation, robust reconstruction and optimal recovery of block models. In *Conference on Learning Theory*, pages 356–370. PMLR, 2014. [3](#)
- [MNS15] Elchanan Mossel, Joe Neeman, and Allan Sly. Reconstruction and estimation in the planted partition model. *Probability Theory and Related Fields*, 162(3-4):431–461, 2015. [1](#), [3](#)
- [MNS18] Elchanan Mossel, Joe Neeman, and Allan Sly. A Proof of the Block Model Threshold Conjecture. *Combinatorica*, 38(3):665–708, Jun 2018. [3](#)
- [Mon98] Rémi Monasson. Optimization problems and replica symmetry breaking in finite connectivity spin glasses. *Journal of Physics A: Mathematical and General*, 31(2):513, 1998. [2](#)

- [Mon23] Andrea Montanari. Sampling, diffusions, and stochastic localization. *arXiv preprint arXiv:2305.10690*, 2023. [4](#)
- [MP01] Marc Mézard and Giorgio Parisi. The Bethe lattice spin glass revisited. *The European Physical Journal B*, 20:217–233, 2001. [2](#)
- [MP03] Marc Mézard and Giorgio Parisi. The Cavity Method at Zero Temperature. *Journal of Statistical Physics*, 111:1–34, 2003. [2](#)
- [MS13] Elchanan Mossel and Allan Sly. Exact thresholds for Ising–Gibbs samplers on general graphs. 2013. [4](#)
- [MW23] Andrea Montanari and Yuchen Wu. Posterior sampling from the spiked models via diffusion processes. *arXiv preprint arXiv:2304.11449*, 2023. [4](#)
- [Roc24] Sebastien Roch. *Modern Discrete Probability: An Essential Toolkit*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2024. [62](#)
- [SM22] Ludovic Stephan and Laurent Massoulié. Non-backtracking spectra of weighted inhomogeneous random graphs. *Mathematical Statistics and Learning*, 5(3):201–271, 2022. [22](#)
- [SS12] Allan Sly and Nike Sun. The Computational Hardness of Counting in Two-Spin Models on d -Regular Graphs. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, pages 361–369, 2012. [1](#)
- [ST96] Harold M Stark and Audrey A Terras. Zeta functions of finite graphs and coverings. *Advances in Mathematics*, 121(1):124–165, 1996. [21](#)
- [Tal11] Michel Talagrand. *Mean Field Models for Spin Glasses: Volume I: Basic Examples*, volume 54 of *A Series of Modern Surveys in Mathematics*. Springer-Verlag, 2011. [1](#)
- [WJ08] Martin J. Wainwright and Michael I. Jordan. *Graphical Models, Exponential Families, and Variational Inference*, volume 1 of *Foundations and Trends in Machine Learning*. Now Publishers Inc, 2008. [1](#)
- [YP23] Qian Yu and Yury Polyanskiy. Ising model on locally tree-like graphs: Uniqueness of solutions to cavity equations. *IEEE Transactions on Information Theory*, 2023. [3](#)

A Proofs omitted from the main text

A.1 Generalized MLSI for Ising models on pseudorandom near-forests

We begin with a proof of [Lemma 4.19](#). While the proof we state in the main text which assumes that all nonzero interactions are equal suffices for our purposes, a more general result holds.

Lemma 4.19. *Let H be a (Δ, D) -pseudorandom tree on n vertices, and J an interaction matrix supported on H with all interactions in $\left[0, \frac{\gamma}{\sqrt{D}}\right]$ for some $\gamma \in [0, 1)$. There is a constant c (depending on γ) such that for any external field $h \in \mathbb{R}^n$,*

$$C_{\text{MLSI}}(\mu_{J,h}) \geq \frac{1}{ne^{c\Delta}}.$$

Proof. To prove this statement, as in the simpler proof given in the main text, we will use [Lemma 3.18](#) and [Theorem 3.19](#), albeit with the stochastic localization using a time-varying control matrix. Again, set $s_0 = \frac{\gamma}{\sqrt{D}}$. Consider the graph H_t on $V(H)$ with edge set

$$E(H_t) = \left\{ ij \in E(H) : J_{ij} > t \cdot \frac{\gamma}{\sqrt{D}} \right\}.$$

In particular, H_1 is empty. The control matrix at time t is defined by $C_t = \left((1 - s_0^2) \cdot \text{BH}_{H_t} \left(-\frac{\gamma}{\sqrt{D}} \right) \right)^{1/2}$. As in the simpler proof, this is a non-negative matrix that is well-defined due to [Corollary 3.24](#). At time t , the interaction matrix J_t is, on the off-diagonal elements, supported on the graph H_t , with the interactions being at most $\frac{\gamma}{\sqrt{D}}$. In particular, J_1 is a diagonal matrix so μ_1 is a product measure. Now, for an arbitrary tilt h ,

$$\mathbf{Cov}(\mu_{J_t,h}) \leq \sum_{\ell \geq 0} \left(\frac{\gamma}{\sqrt{D}} \right)^\ell A_{H_t}^{(\ell)} \leq \sum_{\ell \geq 0} \left(\frac{\gamma}{\sqrt{D}} \right)^\ell A_H^{(\ell)} = \text{BH}_H(s_0)^{-1},$$

where the inequalities mean that the entries of the (non-negative) matrix on the left are at most that of the corresponding entries on the right. Similarly, because H_t is a subgraph of H , $C_t^2 \leq (1 - s_0^2) \text{BH}_H(-s_0)$. Using [Corollary 4.16](#) and [Lemma 4.18](#), we have that

$$\begin{aligned} \|C_t \cdot \mathbf{Cov}(\mu_{J_t,h}) \cdot C_t\| &\leq \|\mathbf{Cov}(\mu_{J_t,h}) \cdot C_t^2\| \\ &\leq (1 - s_0^2) \left\| \text{BH}_H(s_0)^{-1} \cdot \text{BH}_H(-s_0) \right\| \\ &= (1 - s_0^2) \left\| \text{BH}_H(s_0)^{-1} \cdot \left(\text{BH}_H(s_0) + \frac{2s_0}{1 - s_0^2} \cdot A_H \right) \right\| \\ &\leq (1 - s_0^2) + \frac{2s_0}{1 - s_0^2} \left\| \text{BH}_H(s_0)^{-1} \cdot A_H \right\| \\ &\leq 1 + \frac{8}{(1 - \gamma)^2} \cdot \frac{\Delta}{D}. \end{aligned}$$

The measure at time 1 is almost surely a product measure, therefore

$$\begin{aligned} C_{\text{MLSI}}(\mu_{J,h}) &\geq \frac{1}{n} \cdot \exp \left(- \int_0^1 \sup_{h \in \mathbb{R}^n} \|C_t \cdot \mathbf{Cov}(\mu_{J_t,h}) \cdot C_t\| \right) \\ &\geq \frac{1}{n} \cdot \exp \left(- \left(1 + \frac{8}{(1 - \gamma)^2} \cdot \frac{\Delta}{D} \right) \right) = \frac{1}{ne^{O(\Delta)}}. \end{aligned} \quad \blacksquare$$

A.2 Bounds on various probabilities in sparse random graphs

Next, we prove numerous lemmas from [Section 6](#), including [Lemmas 6.4, 6.14, 6.17](#) and [6.19](#). In the sequel, we study random graphs G sampled from $\text{SBM}(n, d, \lambda) | \sigma$ where σ is a “balanced” community assignment. In particular, we assume that $\frac{1}{n} |\langle \mathbf{1}, \sigma \rangle| < O\left(\sqrt{\frac{\log n}{n}}\right)$. This holds with probability 1 if G is an Erdős–Rényi graph, and more generally with probability $1 - o_n(1)$ for 2-community SBMs with nontrivial priors.

Lemma A.1. *For any vertex v in $V(G)$ and $r \geq 1$, for $\eta = \min\left\{\frac{\varepsilon^2}{54}, \frac{1}{18}\right\}$,*

$$\Pr[|B_r(v)| \geq (d(1 + \varepsilon))^r] \leq \exp\left(-\eta d \left(1 + \frac{\varepsilon}{4}\right)^r\right).$$

Proof. Consider a Galton-Watson tree T with branching random variable $\text{Poi}(d(1 + o_n(1)))$. It is well-known (see e.g. [\[Roc24, Example 4.2.8\]](#)) that $\text{Poi}(d)$ stochastically dominates $\text{Bin}(n - 1, \frac{d}{n})$, so one can construct a monotone coupling between the local neighborhoods of G and the Galton-Watson tree T in the following way. Suppose α is the fraction of vertices v such that $\sigma(v) = +1$. Then, the number of neighbors of a given vertex v is distributed as $\text{Bin}\left(\alpha n - 1, \frac{d + \lambda \sqrt{n}}{n}\right) + \text{Bin}\left((1 - \alpha)n, \frac{d - \lambda \sqrt{n}}{n}\right)$, which is dominated by $\text{Poi}\left(\alpha \cdot \frac{d + \lambda \sqrt{d}}{2} + (1 - \alpha) \cdot \frac{d - \lambda \sqrt{d}}{2}\right)$. Since $\alpha = \frac{1}{2} \pm O\left(\sqrt{\frac{\log n}{n}}\right)$, the number of neighbors is stochastically dominated by $\text{Poi}(d(1 + o_n(1)))$. For simplicity of notation, for the rest of the proof we replace $d(1 + o_n(1))$ with d . Since $\varepsilon = \Omega(\log d / d)^{1/3}$, we can increase the implicit constant to ensure that the original tail bound as stated holds.

Let u be the root of T . Set $N_\ell^{\text{GW}}(u)$ to be the set of vertices at depth exactly ℓ in the Galton-Watson tree, and $B_\ell^{\text{GW}}(u)$ to be the set of vertices at depth at most ℓ . By the above monotone coupling, $|B_L(v)|$ is stochastically dominated by $|B_L^{\text{GW}}(u)|$, so it suffices to show the relevant tail bound for $|B_\ell^{\text{GW}}(u)|$.

For ease of notation, set $R_t = |N_t^{\text{GW}}(u)|$ and $B_t = |B_t^{\text{GW}}(u)|$. Also let $B_t = 0$ for $t < 0$. Markov’s inequality implies that for any $s > 0$,

$$\Pr[|B_r| \geq (d(1 + \varepsilon))^r] \leq e^{-s(d(1 + \varepsilon))^r} \mathbf{E}\left[e^{sB_r}\right].$$

Now, for $t \geq 2$, using the recursion $B_t = B_{t-1} + R_t$, we have

$$\begin{aligned} \mathbf{E}\left[e^{sB_t}\right] &= \mathbf{E}\left[e^{sB_{t-1}} \cdot e^{sR_t}\right] \\ &= \mathbf{E}\left[e^{sB_{t-1}} \cdot e^{d(e^s - 1)R_{t-1}}\right] && \text{(Poisson MGF)} \\ &= \mathbf{E}\left[e^{sB_{t-2}} \cdot e^{(d(e^s - 1) + s)R_{t-1}}\right]. \end{aligned}$$

Define $g : \mathbb{R}_{\geq 0}^2 \rightarrow \mathbb{R}_{\geq 0}$ by

$$g(x, s) = d(e^x - 1) + s,$$

so for any $r \geq 1$, $\mathbf{E}[e^{sB_r}] \leq e^{g^{\circ r}(s, s)}$, where $g^{\circ 1}(x, y) = g(x, y)$ and $g^{\circ(t+1)}(x, y) = g(g^{\circ t}(x, y), y)$. Now, define $s_k = \delta(d + 2\delta)^{-k}$ for some $\delta \geq 1/d$ that we shall set later. Then, for any $1 \leq t \leq r$,

$$g(s_t, s_r) = d(e^{s_t} - 1) + s_r$$

$$\begin{aligned}
&\leq ds_t e^{s_t} + s_r \\
&\leq s_{t-1} \cdot \left(\frac{de^{s_t}}{d+2\delta} + \frac{s_r}{s_{t-1}} \right) \\
&\leq s_{t-1} \cdot \left(\frac{de^{s_t}}{d+2\delta} + \frac{1}{d+2\delta} \right) \\
&\leq s_{t-1} \cdot \left(\frac{d \exp\left(\frac{\delta}{d+2\delta}\right)}{d+2\delta} + \frac{d}{d+2\delta} \right) \quad (t \geq 1) \\
&\leq s_{t-1}.
\end{aligned}$$

Here, the final inequality follows from the fact that

$$\frac{\exp\left(\frac{x}{1+2x}\right)}{1+2x} + \frac{1}{1+2x} \leq 1$$

for $x \geq 1$, which we have applied with $x = \frac{\delta}{d}$. We also have

$$g(s_0, s_r) = d(e^{s_0} - 1) + s_r \leq d\delta e^\delta + s_r \leq d\delta e^\delta + \delta.$$

Consequently,

$$\mathbf{E} \left[e^{s_{r-1} B_r} \right] \leq e^{d\delta e^\delta + \delta}.$$

Therefore,

$$\begin{aligned}
\Pr [|B_r| \geq (d(1+\varepsilon))^r] &\leq e^{d\delta e^\delta + \delta} \cdot e^{-s_{r-1}(d(1+\varepsilon))^r} \\
&= \exp(d\delta e^\delta + \delta) \cdot \exp\left(-\delta \cdot \frac{(d(1+\varepsilon))^r}{(d+2\delta)^{r-1}}\right) \\
&\leq \exp(d\delta e^\delta + \delta) \cdot \exp\left(-\delta d \cdot \left(\frac{1+\varepsilon}{1+\frac{2\delta}{d}}\right)^r\right).
\end{aligned}$$

Now, choose $\delta = \min\left\{\frac{\varepsilon}{4+2\varepsilon}, \log\left(1+\frac{\varepsilon}{8}\right)\right\}$. Recall that $\varepsilon = \Omega\left(\frac{\log d}{d}\right)^{1/3}$, so this is indeed at least $1/d$. This implies that $\frac{1+\varepsilon}{1+\frac{2\delta}{d}} \geq 1 + \frac{\varepsilon}{2}$. Furthermore,

$$\begin{aligned}
\frac{\exp(d\delta e^\delta + \delta) \cdot \exp\left(-\delta d \left(1 + \frac{\varepsilon}{2}\right)^r\right)}{\exp\left(-\frac{\delta \varepsilon d}{4+\varepsilon} \left(1 + \frac{\varepsilon}{4}\right)^r\right)} &= \exp\left(d\delta e^\delta + \delta - d\delta \left(\left(1 + \frac{\varepsilon}{2}\right)^r - \frac{\varepsilon}{4+\varepsilon} \left(1 + \frac{\varepsilon}{4}\right)^r\right)\right) \\
&\leq \exp\left(d\delta e^\delta + \delta - d\delta \left(1 + \frac{\varepsilon}{2} - \frac{\varepsilon}{4+\varepsilon} \left(1 + \frac{\varepsilon}{4}\right)\right)\right) \\
&= \exp\left(d\delta e^\delta + \delta - d\delta \left(1 + \frac{\varepsilon}{4}\right)\right) \\
&= \exp\left(d\delta \left(e^\delta + \frac{1}{d} - \left(1 + \frac{\varepsilon}{4}\right)\right)\right) \\
&\leq \exp\left(d\delta \left(e^\delta - \left(1 + \frac{\varepsilon}{8}\right)\right)\right) \leq 1,
\end{aligned}$$

where in the last line we used $\varepsilon \geq \frac{8}{d}$. To conclude, we set

$$\begin{aligned}\eta &:= \frac{\delta\varepsilon}{4+\varepsilon} = \frac{\varepsilon}{4+\varepsilon} \cdot \min \left\{ \frac{\varepsilon}{4+2\varepsilon}, \log \left(1 + \frac{\varepsilon}{8} \right) \right\} \\ &\geq \min \left\{ \frac{\varepsilon}{6}, \frac{1}{3} \right\} \cdot \min \left\{ \frac{\varepsilon}{9}, \frac{1}{6} \right\} \\ &\geq \min \left\{ \frac{\varepsilon^2}{54}, \frac{1}{18} \right\}.\end{aligned}$$

First, observe that [Lemma 6.14](#), restated below, is an immediate corollary of the above. ■

Lemma 6.14. *For $\kappa = \frac{4}{\varepsilon} \log \frac{\log n}{\eta d}$, we have:*

$$\Pr[\exists v \in [n] \text{ s.t. } \ell(v) > \kappa] = o(1).$$

Now, we proceed to fill in other missing proofs.

Lemma 6.4. *With probability $1 - o(1)$, for every vertex v in G and $\ell \geq 0$, $\frac{|B_\ell(v)|}{(d(1+\varepsilon))^\ell} = o(\log n)$.*

Proof. As in the proof of [Lemma A.1](#), set R_ℓ to be the number of vertices at depth ℓ in a Galton-Watson tree with branching random variable $\text{Poi}(d)$ rooted at v . We shall show that

$$\Pr \left[R_\ell \geq (d(1+\varepsilon))^\ell \cdot \Delta \right] \leq \frac{1}{n^3}$$

for some $\Delta = o(\log n)$ (Δ is $o(\log n)$ uniformly for all ℓ). Given this, we are done by taking a union bound over $0 \leq \ell \leq n$ and then all n vertices v . Again, define

$$g(x) = d(e^x - 1),$$

so for any $s > 0$,

$$\Pr \left[R_\ell \geq (d(1+\varepsilon))^\ell \cdot \Delta \right] \leq e^{-s\Delta(d(1+\varepsilon))^\ell} \cdot e^{s\Delta(d(1+\varepsilon))^\ell}.$$

Alternately, set

$$h(x) = g^{-1}(x) = \log \left(\frac{x}{d} + 1 \right),$$

so we get that for any $q > 0$,

$$\Pr \left[R_\ell \geq (d(1+\varepsilon))^\ell \cdot \Delta \right] \leq e^{-h^{\circ\ell}(q)\Delta(d(1+\varepsilon))^\ell} \cdot e^q.$$

Now, set $q = \log n$ and

$$\Delta = \frac{4 \log n}{(d(1+\varepsilon))^\ell \cdot h^{\circ\ell}(\log n)}.$$

Clearly, this yields the $1/n^3$ bound claimed earlier. Further, we have

$$\min_{\ell \geq 0} \left\{ (d(1+\varepsilon))^\ell \cdot h^{\circ\ell}(\log n) \right\} = \omega(1),$$

so $\Delta = o(\log n)$. ■

The key idea in the proof of [Lemma 6.19](#) is the following observation.

Observation A.2. Let $S \subseteq V$ be any set of vertices in G and ℓ be a positive integer. For any vertex $v \in V$ and conditioning on the induced subgraph on S , we have that $N_\ell(v) \setminus S$ is a uniformly random subset of $V \setminus S$ of size $|N_\ell(v) \setminus S|$.

Lemma 6.19. Let S, T_1, T be disjoint subsets of vertices, and say $T = T_2 \sqcup T'$. Let \mathcal{E} be an (arbitrary) conditioning of the edges adjacent to vertices in S, T_1, T such that for any $A = S, T_1, T$, we can write $A = A_{\text{in}} \sqcup A_{\text{out}}$ such that the following holds.

- $|A| = o(n)$.
- Every edge incident on vertices in A_{in} is fixed in a way that there are no edges between A_{out} and A^c .
- Further, there is no conditioning done on the edges between A_{out} and $(S_{\text{in}} \cup (T_1)_{\text{in}} \cup T_{\text{in}})^c$.

Let $u \in [n]$ and $\ell > 0$. Then, for n sufficiently large,

$$\Pr[u \text{ is } \ell\text{-light and } B_\ell(u) \cap T_i \neq \emptyset \text{ for } i \in \{1, 2\} \mid \mathcal{E}] \leq \frac{2\ell^2 \cdot (d(1 + \varepsilon))^{2\ell} \cdot |T_1| \cdot |T_2|}{n^2}.$$

Proof of Lemma 6.19. The idea is that due to [Observation A.2](#) — since the sets S, T_1, T are all $o(n)$ in size, we are essentially picking uniformly random small subsets of vertices on V and checking whether they intersect.

To formalize this, consider growing out the ball around u incrementally. Let us first introduce some additional notation. Set \mathcal{L} to be the event that u is ℓ -light, and let \mathcal{F} be the event in the lemma statement whose conditional probability we wish to bound:

$$\mathcal{F} := \mathcal{L} \cap \{B_\ell(u) \cap T_i \neq \emptyset \text{ for } i \in \{1, 2\}\}$$

For $r \geq 1$ and $i \in [2]$, let $\mathcal{E}_{r,i}$ denote the event that $B_r(u) \cap T_i \neq \emptyset$ but $B_{r-1}(u) \cap T_i = \emptyset$, that is, r is the smallest radius at which the ball intersects T_i . Noting that $\mathcal{E}_{r,i}$ are disjoint events for fixed i , we have

$$\Pr[\mathcal{F} \mid \mathcal{E}] = \sum_{r_1, r_2=1}^{\ell} \Pr[\mathcal{L} \cap \mathcal{E}_{r_1,1} \cap \mathcal{E}_{r_2,2} \mid \mathcal{E}]. \quad (16)$$

In words, we are splitting this into cases depending on when T_1 and T_2 are hit. To complete the proof, we shall bound each of the terms in this summation by

$$\frac{2(d(1 + \varepsilon))^{2\ell} \cdot |T_1| \cdot |T_2|}{n^2}.$$

Fix r_1, r_2 with $r_1 \leq r_2$. Let \mathcal{L}_r be the event that $|B_r(u)| \leq (d(1 + \varepsilon))^\ell$. We have

$$\begin{aligned} \Pr[\mathcal{E}_{r_1,1} \cap \mathcal{E}_{r_2,2} \cap \mathcal{L} \mid \mathcal{E}] &\leq \Pr[\mathcal{E}_{r_1,1} \cap \mathcal{E}_{r_2,2} \cap \mathcal{L}_{r_2} \mid \mathcal{E}] \\ &\leq \Pr[\mathcal{E}_{r_1,1} \cap \mathcal{L}_{r_2} \mid \mathcal{E}] \cdot \Pr[\mathcal{E}_{r_2,2} \mid \mathcal{L}_{r_2} \cap \mathcal{E}_{r_1,1} \cap \mathcal{E}] \\ &\leq \Pr[\mathcal{E}_{r_1,1} \mid \mathcal{L}_{r_1} \cap \mathcal{E}] \cdot \Pr[\mathcal{E}_{r_2,2} \mid \mathcal{L}_{r_2} \cap \mathcal{E}_{r_1,1} \cap \mathcal{E}]. \end{aligned}$$

Let us start by bounding the first of these two probabilities; refer to [Figure 4](#) for a depiction of the sets involved. For convenience, we will denote $V' = V \setminus (T_2 \cup S)$. Suppose that $B_{r_1-1}(u)$, $|N_{r_1}(u)|$ and $|N_{r_1}(u) \cap V'|$ were fixed in a way that $B_{r_1-1}(u)$ does not intersect T_1 or T' , and \mathcal{L}_{r_1} is satisfied. Then, $N_{r_1}(u) \cap V'$ is in fact a uniformly random subset of $V' \setminus B_{r_1-1}(u)$ of the appropriate size. Since T_1 is assumed to be contained in V' , it follows that

$$\Pr[\mathcal{E}_{r_1,1} \mid \mathcal{L}_{r_1} \cap \mathcal{E}] \leq \frac{\mathbf{E} \left[|N_{r_1}(u)| \mid \mathcal{L}_{r_1} \cap \mathcal{E} \right] \cdot |T_1|}{n - o(n)} \leq \frac{2(d(1+\varepsilon))^\ell \cdot |T_1|}{n}.$$

Similarly, we also get that

$$\Pr[\mathcal{E}_{r_2,2} \mid \mathcal{L}_{r_2} \cap \mathcal{E}_{r_1,1} \cap \mathcal{E}] \leq \frac{2(d(1+\varepsilon))^\ell \cdot |T_2|}{n}.$$

Plugging these two bounds into [\(16\)](#), we get that

$$\Pr[\mathcal{F} \mid \mathcal{E}] \leq \frac{2\ell^2 \cdot (d(1+\varepsilon))^{2\ell} \cdot |T_1| \cdot |T_2|}{n^2},$$

as claimed. ■

Lemma 6.17. *Let $u_1, u_2, \dots, u_r \in [n]$ and $\ell_1, \ell_2, \dots, \ell_r > 0$. Then,*

$$\Pr \left[\ell(u_i) = \ell_i \ \forall i \in [r] \text{ and } B_{\ell_i}(u_i) \cap B_{\ell_j}(u_j) = \emptyset \ \forall i \neq j \in [r] \right] \leq \prod_{i \in [r]} \exp \left(-\eta d \left(1 + \frac{\varepsilon}{4} \right)^{\ell_i} \right).$$

The key observation for this proof is the following.

Observation A.3. Let $S \subseteq V$ be a (small) subset of vertices in G . Consider a partial conditioning of the edges for which

- the edges within S are arbitrarily conditioned on.
- for any potential edge e from S to S^c , either e is not conditioned on at all, or is conditioned to be absent.

Now, fix some vertex u , and suppose we grow out a radius r ball from v , conditioned on it not intersecting S . Then, this is equivalent to growing out a radius r ball from v within a smaller completely random graph on $V \setminus S$, and then adding back the edges corresponding to S , with a further conditioning that there are no edges between $B_{r-1}(u)$, the interior vertices of the ball, and $V \setminus B_r(u)$ (which, in particular, includes S).

Consequently, any monotonically nondecreasing function of the sizes of neighborhoods in the conditioned graph are stochastically dominated by their versions without conditioning. In particular, the random variable $\ell(v)$ in the conditioned graph is stochastically dominated by $\ell(v)$ in $G(n, d/n)$.

Proof of Lemma 6.17. We will upper bound the probability inductively; the base case immediately follows from Lemma A.1.

Let \mathcal{A}_r denote the event that $\ell(u_i) = \ell_i$ for every $i \in [r]$, i.e. the first r balls are heavy. Further, let \mathcal{B}_r denote the event that $B_{\ell_i}(u_i) \cap B_{\ell_j}(u_j) = \emptyset$ for all distinct $i, j \in [r]$, i.e. the first r balls are pairwise disjoint. Noting that $\mathcal{B}_r = \mathcal{B}_{r-1} \cap \{B_{\ell_r}(u_r) \cap B_{\ell_i}(u_i) = \emptyset \forall i \in [r-1]\}$, we have

$$\begin{aligned} \Pr[\mathcal{A}_r \cap \mathcal{B}_r] &= \Pr[\mathcal{A}_{r-1} \cap \mathcal{B}_{r-1}] \cdot \Pr\left[B_{\ell_r}(u_r) \cap B_{\ell_i}(u_i) = \emptyset \forall i \in [r-1] \mid \mathcal{A}_{r-1} \cap \mathcal{B}_{r-1}\right] \\ &\quad \cdot \Pr\left[\ell(u_r) = \ell_r \mid \mathcal{A}_{r-1} \cap \mathcal{B}_r\right] \\ &\leq \prod_{i \in [r-1]} \exp(-\eta d(1+\varepsilon)^{\ell_i}) \cdot \Pr\left[\ell(u_r) = \ell_r \mid \mathcal{A}_{r-1} \cap \mathcal{B}_r\right] \\ &\leq \prod_{i \in [r]} \exp(-\eta d(1+\varepsilon)^{\ell_i}), \end{aligned}$$

where in the last line we have used Observation A.3 and Lemma A.1 with $S = \bigcup_{i=1}^{r-1} B_{\ell_i}(u_i)$. ■