

METASTABLE MIXING OF MARKOV CHAINS: EFFICIENTLY SAMPLING LOW TEMPERATURE EXPONENTIAL RANDOM GRAPHS

BY GUY BRESLER^{1,a}, DHEERAJ NAGARAJ^{2,b}, AND ESHAAN NICHANI^{3,c}

¹*Department of Electrical Engineering and Computer Science, MIT, [a guy@mit.edu](mailto:guy@mit.edu)*

²*Google Research, [b dheerajnagaraj@google.com](mailto:dheerajnagaraj@google.com)*

³*Department of Electrical and Computer Engineering, Princeton University, [c eshnich@princeton.edu](mailto:eshnich@princeton.edu)*

In this paper we consider the problem of sampling from the low-temperature exponential random graph model (ERGM). The usual approach is via Markov chain Monte Carlo, but Bhamidi et al. showed that any local Markov chain suffers from an exponentially large mixing time due to metastable states. We instead consider *metastable mixing*, a notion of approximate mixing relative to the stationary distribution, for which it turns out to suffice to mix only within a collection of metastable states. We show that the Glauber dynamics for the ERGM at any temperature – except at a lower-dimensional critical set of parameters – when initialized at $G(n, p)$ for the right choice of p has a metastable mixing time of $O(n^2 \log n)$ to within total variation distance $\exp(-\Omega(n))$.

1. Introduction. Given a vector of real-valued parameters $\beta := (\beta_0, \beta_1, \dots, \beta_K) \in \mathbb{R} \times (\mathbb{R}^+)^K$, the exponential random graph model $\text{ERGM}(n, \beta)$ is defined to be the probability measure over all simple graphs with n vertices

$$\mu_\beta(X) = \frac{1}{Z_\beta} \exp\left(\mathcal{H}_\beta(X)\right) \quad ; \quad \mathcal{H}_\beta(X) := \sum_{i=0}^K n^2 \beta_i N_i(X) .$$

The N_i are homomorphism densities corresponding to fixed finite and connected graphs G_0, G_1, \dots, G_K (such as edges, triangles, 4-cycles, 5-cycles, 2-stars, etc, with G_0 being fixed to be a single edge connecting two vertices) and Z_β denotes the normalizing constant. This is an exponential family where the sufficient statistics are the subgraph counts. The model significantly generalizes the Erdős-Rényi random graph and is used to model a variety of complex networks like social networks and biological networks [10, 16, 8, 9, 31]. Early analysis was carried out by statistical physicists [24, 25, 4], and probabilists and statisticians have further studied various questions about these models including sampling, estimation, large deviations theory, concentration of measure, and phase transitions [1, 5, 26, 7, 11, 23, 32, 27].

The basic problem we consider in this work is that of producing a sample from the ERGM probability distribution in polynomial time. A popular approach to sampling is to use the Glauber dynamics, a simple reversible Markov chain with the desired stationary distribution, and to run it for sufficiently long that it is close to stationarity.

DEFINITION 1.1 (Glauber Dynamics). Given any probability distribution π over $\mathcal{X} = \{0, 1\}^{\binom{n}{2}}$ with $\pi(x) > 0$ for all $x \in \mathcal{X}$, the Glauber dynamics with respect to π is the discrete time Markov chain over \mathcal{X} with single step transition from X to X' as follows:

1. Pick a coordinate $I \in \left[\binom{n}{2} \right]$ uniformly at random.

MSC2020 subject classifications: Primary 60J10; secondary 68Q87.

Keywords and phrases: Exponential Random Graphs, Efficient Sampling.

2. Form X' by keeping all edges except for I the same as in X and sample $X'_I \sim \pi(\cdot | X_{\sim I})$ conditional on the other edges. Here $X_{\sim I}$ denotes all the edges of X other than the edge I .

The Glauber dynamics with respect to π is reversible for π ([20]), so in particular π is stationary.

The mixing time of the Glauber dynamics, i.e., the time until the distribution is within total variation $1/4$ of stationarity, determines whether the approach is feasible. This mixing time was shown by Bhamidi et al. [1] to be essentially characterized by the function $L_\beta : [0, 1] \rightarrow \mathbb{R}$ defined as

$$(1) \quad L_\beta(p) = \sum_{i=0}^K \beta_i p^{|E_i|} - I(p)$$

where for each i , E_i is the edge set of the fixed graph G_i and $I(p) := \frac{1}{2}p \log p + \frac{1}{2}(1-p) \log(1-p)$. (Bhamidi et al. [1] actually formulated their results in an equivalent way in terms of the function ϕ_β defined in Section 2.3, while L_β was studied by [6].)

THEOREM 1.2 ([1]). *Consider the ERGM(n, β) distribution. There are three regimes for β :*

1. *High temperature: If L_β has a unique local maximum with a non-vanishing second derivative, then the Glauber dynamics Markov chain for ERGM(n, β) mixes in time $O(n^2 \log n)$.*
2. *Low temperature: If L_β has multiple local maxima with non-vanishing second derivatives, then any local Markov chain with stationary distribution ERGM(n, β) must suffer a mixing time of $\exp(\Omega(n))$.*
3. *Critical temperature: If any local maxima of L_β has a vanishing second derivative.*

In this work, we use a somewhat informal notion of metastability since this is not needed in order to state our technical results. Consider subsets of the state space $A, B \subseteq \mathcal{X}$ such that $A \subseteq B$. We will call B to be metastable with respect to a given Markov dynamics if the Markov dynamics initialized inside the set A takes a long time – exponential in n – to exit the set B . Slow mixing in the low-temperature phase is due to the existence of multiple, disconnected metastable states from which it takes the Glauber dynamics exponential time to leave. The question is therefore: can one efficiently produce a sample from the ERGM in the low temperature regime?

An important insight into the structure of the ERGM distribution was developed by Chatterjee and Diaconis [6], and this will constitute a useful step towards our goal. They showed that the ERGM distribution is close to a finite mixture of constant graphons with respect to the cut-metric (graphons and the cut metric δ_\square are reviewed in Section 2.2):

THEOREM 1.3 (Theorem 4.2 of [6]). *Denote by M_β the set of global maxima of L_β . The ERGM(n, β) distribution converges, in probability with respect to the cut-metric, to a mixture of $G(n, p^*)$ for $p^* \in M_\beta$. Formally, let $X_n \sim \text{ERGM}(n, \beta)$ and \tilde{X}_n be its corresponding graphon and \tilde{M}_β be the set of all constant graphons with value p^* for some $p^* \in M_\beta$. For every fixed $\eta > 0$, there are constants $C(\eta), c(\eta) > 0$ such that*

$$\mathbb{P}(\delta_\square(\tilde{X}_n, \tilde{M}_\beta) > \eta) \leq C(\eta) \exp(-c(\eta)n^2).$$

Given the approximation results of Chatterjee and Diaconis, can one simply find p^* and obtain a sufficiently accurate approximation of the ERGM by sampling $G(n, p^*)$? Unfortunately, no: Theorem 1.3 is in *cut metric*, which turns out to be too weak to control total

variation distance. Indeed, the paper [3] shows that $d_{\text{TV}}(\text{ERGM}, G(n, p^*)) \rightarrow 1$ even for the simple ERGM consisting of edges and 2-stars. The approximation results of Eldan and Gross [7] show that even in the low-temperature regime, the ERGM can be approximated in a certain Wasserstein metric by an appropriate mixture of stochastic block models, endowing it with richer structure compared to $G(n, p^*)$. However, these results also do not imply approximation in the total variation distance.

The ERGM is closely related to the ferromagnetic Ising model and, in fact, the ERGM with the 2-star can be written as an Ising model. Just as for the ERGM, the Glauber dynamics is known to mix exponentially slowly for the Ising model at low temperatures. Nevertheless, there are Markov chains mixing in polynomial-time which can sample efficiently from arbitrary ferromagnetic Ising models based on random cluster dynamics [15] and the Swendsen-Wang dynamics [29, 30]. For the ERGM it is not at all clear how to write down the corresponding random cluster model such that the associated random cluster dynamics mixes rapidly. We instead pursue a more direct approach.

The starting point of our approach is the observation that if our aim is only to produce a sample from nearly the correct distribution, then there is no need for the dynamics to transition between all metastable states. In order to implement this intuition, it is necessary to slightly modify the standard definition of mixing time of a Markov chain. The usual definition measures the distance to stationarity starting from a worst-possible initial state. Instead, we use the following definition.

DEFINITION 1.4. Given a Markov transition kernel P with stationary π^* we start from some initial distribution π_0 and say that P is $(\pi_0, \pi^*, \tau, \delta)$ -**mixing** if for every $t \geq \tau$

$$(2) \quad d_{\text{TV}}(\pi_0 P^t, \pi^*) \leq \delta.$$

Note that an immediate consequence of the data processing inequality for total variation is that P is $(\pi_0, \pi^*, \tau, \delta)$ -mixing if and only if $d_{\text{TV}}(\pi_0 P^\tau, \pi^*) \leq \delta$.

Our main result, Theorem 3.1 given in Section 3, shows that whenever $\delta \geq \delta_0 = \exp(-\Omega_\beta(n))$, with π_0 being a mixture of $G(n, p^*)$ for some carefully chosen distribution p^* , the Glauber dynamics for ERGM is $(\pi_0, \pi^*, C_\beta n^2 \log \frac{n}{\delta}, \delta)$ -mixing even in the low temperature regime. That is, as long as the target TV distance is $\geq \delta_0$, then the mixing time of the Glauber dynamics is $O(n^2 \log n)$. This gives a counterpoint to the criticism of these models in [1] based on the difficulty of sampling these models at low temperature.

The role of δ_0 merits some discussion. Incorporating the starting distribution into the definition of mixing time invalidates one of the basic lemmas: it is no longer true that the total variation decreases exponentially as $\exp(-Ct/n^2 \log(n))$ for every t . The basic reason is that in a Markov chain with multiple metastable states requiring exponential time to leave, any initial error in probability assigned to the metastable states might persist for exponential time. Thus, one might think of δ_0 as capturing this initial (possibly unavoidable) error. We remark that Gheissari and Sinclair's work [14] on mixing in low-temperature Ising models also considers mixing up to a TV distance of δ_0 .

The following is a corollary of Theorem 3.1.

THEOREM 1.5. Suppose that p^* is the unique global maximizer of L_β and moreover that L_β has nonzero second derivative at p^* . Let $\pi_0 := G(n, p^*)$. There exist positive constants c_β, C_β and $n_0(\beta)$, such that if $n > n_0(\beta)$, then whenever $\delta \geq \exp(-c_\beta n)$, the Glauber dynamics for μ_β is $(\pi_0, \mu, C_\beta n^2 \log(n^2/\delta), \delta)$ -mixing.

In fact, Theorem 3.1 is more comprehensive and shows that even when there are multiple global maximizers in the low temperature regime, we can sample efficiently from the conditional distribution of being close to any of these maximizers. In Theorem 3.3, we establish

a surprising richness which can be present in the ERGM at low temperature. Even within a small cut-metric neighborhood of the constant p^* graphon where the ERGM measure concentrates (and looks very close to $G(n, p^*)$) [6], we establish the existence of metastable states for certain choices of β whose total probability is $\exp(-\Theta(n))$. The Glauber dynamics takes an exponentially long time to escape this set of metastable states. In contrast, Bhamidi et al. [1] constructed metastable states as sets of graphs similar to $G(n, p)$, where p was a local maximizer of L_β , and these have a total probability of $\exp(-\Omega(n^2))$.

1.1. Overview and Proof Sketch. We now give a high level overview of the ideas behind the main results, which are stated in Section 3.

Sufficient Conditions for Path Coupling. The large deviations results in [6] stated here as Theorem 1.3 show that a sample from the ERGM is w.h.p. close in cut-metric to some constant graphon with value p^* . Sufficient conditions established in [1] for the path coupling argument to work requires much stronger control on small subgraph counts than provided by the cut-metric: the increase in homomorphism density of G in X formed by adding any edge e to X (denoted by $\Delta_G^e(X)$) must be approximately $2n^{-2}|E(G)|(p^*)^{|E(G)|-1}$ for every fixed subgraph G . We will show that $\Delta_G^e(X)$ indeed concentrates close to this value with probability $1 - \exp(-\Omega_\beta(n))$. The coupling argument showing how this statement implies our main theorem is contained in Sections 4 and 5.

Fixed Point Equations for Subgraph Concentration. Theorem 1.3 shows that $X \sim \mu$ is close to the constant graphon with value p^* . Section 6 reduces the task of showing concentration of $\Delta_G^e(X)$ to show additionally that: (1) every node degree uniformly concentrates close to p^* and (2) that the number of common neighbors of any two vertices u and v is close to $n(p^*)^2$. We show these two properties as follows. First, concentration of degrees is established using the cavity method, discussed momentarily. For the second property, we make use of the concentration of degrees to derive a fixed point equation for the common neighbor counts and use the concentration results for fixed point equations established in [5, Theorem 1.5]. We note that the concentration results given by [5, Theorem 1.5] in themselves do not seem to be sufficient to establish the concentration of $\Delta_G^e(X)$ and the concentration of degrees as established by the cavity method is essential.

Cavity Method for Degrees. We use the cavity method as developed in Section 8 to first show that conditioned on the exponential random graph X being close to the constant p^* graphon, the normalized degree of *every* vertex concentrates close to p^* . Graphon convergence can show that *most* vertices have degree close to p^* (see Lemma 7.1). To obtain the uniform concentration, we look at the law of the edges emanating from a single vertex (called the cavity) conditioned on the rest of the graph being close to the constant graphon p^* . We show that the “mean field” generated by the rest of the graph forces the cavity vertex to have degree close to p^* with high probability. This is established in Theorem 8.9 and Corollary 6.3 and is our main technical innovation.

1.2. Discussion and Future Work. Fast mixing of a Markov chain can be used to establish concentration of measure, central limit theorems, and estimation of the partition function. Concentration of measure, CLTs and approximation by $G(n, p)$ ([11, 27]) have been explored in the literature for high-temperature ERGM models. It would be interesting to consider their extension to low-temperature ERGM via the approximate mixing established in this work. Maximum likelihood estimation often involves estimation of the partition function. There are multiple works ([17, 28, 12, 13, 18]) that efficiently approximate partition functions of a parametric family with a given parameter β by efficiently generating samples from the distribution for every choice of the parameter β' . We leave open the problem of estimating the partition function of the low-temperature ERGM.

1.3. *Related Work.* Polynomial or quasipolynomial time mixing from a well-chosen initial distribution has been explored for the mean field Ising model by Levin et al. [19] and more recently for the Ising model on the lattice \mathbb{Z}^d by Gheissari and Sinclair [14]. Lubetzky and Sly [22] consider mixing from specific initial conditions for the 1-dimensional Ising model and identify initial states which allow faster mixing than the worst case by constant factors. The work [21] considers an idea similar to ours in isolating the ‘modes’ of a probability distribution in order to aid sampling. Their approach consider multiple, coupled, random instantiations of the Markov chains, which are all allowed to interact as they evolve, whereas our work considers a single instance of such a Markov chain.

2. Background and Notation. This section contains the basic definitions and notation for graphons and the Glauber dynamics.

2.1. *Notation.* For any simple graph X , we denote by $V(X)$ its vertex set and by $E(X)$ its edge set. Given an unordered pair $e = (u, v)$ for some $u, v \in V(X)$, we define $X_e = X_{uv} = X_{vu} = \mathbf{1}(e \in E(X))$. Without loss of generality, we take $V(X) = [n]$, where $|V(X)| = n$, and identify the space of finite simple graphs on n vertices with the space $\mathcal{X} := \{0, 1\}^{\binom{n}{2}}$, where the coordinates are indexed by tuples (u, v) for $u < v$, $u, v \in [n]$. Throughout, we will reserve u, v, w to denote vertices of size n random graphs for large n and i, j, l to denote vertices of fixed graphs like G_0, G_1, \dots above.

By $X_{\sim e}$ we denote the graph formed by all edges other than the edge e . Given $X \in \mathcal{X}$, define $X^{+e} \in \mathcal{X}$ (resp. $X^{-e} \in \mathcal{X}$) by $(X^{+e})_{\sim e} = X_{\sim e}$ (resp. $(X^{-e})_{\sim e} = X_{\sim e}$) and $(X^{+e})_e = 1$ (resp. $(X^{-e})_e = 0$) i.e., we add (resp. remove) edge e to the graph X .

We use the standard asymptotic notation $O(\cdot)$, $\Omega(\cdot)$, and $\Theta(\cdot)$. For $x, y \in \mathbb{R}^+$ $y = O_\gamma(x)$, we mean $y \leq C_\gamma x$ for some constant C_γ which depends only on γ (and similarly for Ω_γ and Θ_γ). In the statement of the results, expressions of the form $\epsilon < c(\gamma)$ mean “ ϵ smaller than a constant depending only on γ ” and $n > n_0(\gamma)$ means “ n larger than a constant depending only on γ ”.

We will occasionally use the function $I : [0, 1] \rightarrow [-\frac{1}{2}, 0]$ given by $I(p) := \frac{1}{2}p \log p + \frac{1}{2}(1-p) \log(1-p)$. This is just $-1/2$ times the binary entropy function. For any set $A \subseteq B$, we let $A^c := B \setminus A$ whenever the set B is clear from context.

2.2. *Graphon Theory.* Our work relies heavily on the theory of graph limits and graphons; our notation follows [6]. Let \mathcal{W} denote the space of symmetric measurable functions $f : [0, 1]^2 \rightarrow [0, 1]$, where the space $[0, 1]^2$ is endowed with the uniform probability measure. For $f, g \in \mathcal{W}$, define their *cut distance* to be

$$\delta_{\square}(f, g) = \sup_{S, T \subseteq [0, 1]} \left| \int_{S \times T} (f(x, y) - g(x, y)) dx dy \right|,$$

where the supremum is over Borel measurable sets S, T . Define the equivalence relation \sim on \mathcal{W} by $f \sim g$ iff there exists a measure preserving bijection $\sigma : [0, 1] \rightarrow [0, 1]$ such that $f(x, y) = g(\sigma x, \sigma y) := g_\sigma(x, y)$. Let $\tilde{\mathcal{W}}$ be the quotient space with respect to this equivalence relation. For $f \in \mathcal{W}$, let \tilde{f} denote its orbit in \mathcal{W} . A metric δ_{\square} on $\tilde{\mathcal{W}}$ can now be defined as

$$\delta_{\square}(\tilde{f}, \tilde{g}) = \inf_{\sigma} \delta_{\square}(f, g_\sigma).$$

An important fact in the theory of graph limits is that $(\tilde{\mathcal{W}}, \delta_{\square})$ is a compact metric space.

For a graph X with vertex set $[n]$, we can associate the function $f^X \in \mathcal{W}$ where $f^X(x, y) = \mathbf{1}_{(\lfloor nx \rfloor, \lfloor ny \rfloor) \in E(X)} = X_{\lfloor nx \rfloor, \lfloor ny \rfloor}$. We define its corresponding graphon to be

$\tilde{X} = \tilde{f}^X \in \tilde{\mathcal{W}}$. Note that under this mapping, vertex isomorphic graphs correspond to the same element of $\tilde{\mathcal{W}}$. We will denote the graphon with the constant value $p \in [0, 1]$ by $p\mathbf{1}$.

For a finite simple graph H with vertex set $[k]$ for some $k \in \mathbb{N}$ and a graphon $h \in \tilde{W}$, we define the homomorphism density $t(H, h)$ as

$$t(H, h) = \int_{[0,1]^k} \prod_{(i,j) \in E(H)} h(x_i, x_j) dx_1 \dots dx_k.$$

In particular, the subgraph counts $N_i(X)$ appearing in the Hamiltonian $\mathcal{H}_\beta(X)$ are defined as the homomorphism densities

$$(3) \quad N_i(X) := t(G_i, \tilde{f}^X).$$

When emphasizing a particular graph G we will also use the notation $N_G(X)$.

The classical theory of graph limits is too coarse to understand convergence of Markov chains because the cut metric does not control degrees of individual vertices (or neighborhoods of two vertices), which can have a large impact on the evolution of the Glauber dynamics. The following quantities will allow us to establish a fine-grained understanding of the measure μ . Given a graph X with n vertices, whenever $u \in [n]$, we define the normalized degree

$$(4) \quad p_u(X) := \frac{\text{degree of vertex } u}{n} = n \int_0^1 \int_{\frac{u-1}{n}}^{\frac{u}{n}} f^X(x, y) dx dy.$$

Similarly, define the normalized wedge count for nodes u, v by

$$(5) \quad \begin{aligned} p_{uv}(X) &:= \frac{\text{number of vertices to which both } u \text{ and } v \text{ have edges}}{n} \\ &= n^2 \int_0^1 \int_{\frac{u-1}{n}}^{\frac{u}{n}} \int_{\frac{v-1}{n}}^{\frac{v}{n}} f^X(x, z) f^X(y, z) dx dy dz. \end{aligned}$$

2.3. Glauber Dynamics for the ERGM. As described in the introduction, at each step of the Glauber dynamics for μ a pair of vertices $e = \{u, v\}$ is chosen uniformly at random from the $\binom{n}{2}$ possibilities and the variable X_e indicating presence of edge e is updated according to the conditional probability

$$(6) \quad \phi_e(X_{\sim e}) := \mathbb{E}_{X \sim \mu} [X_e | X_{\sim e}].$$

It will be useful to express the update probability in terms of subgraph counts. For any graph $G = (V, E)$ define

$$(7) \quad \Delta_G^e(X) = N_G(X^{+e}) - N_G(X^{-e})$$

and let

$$r_G(X, e) := \left(\frac{n^2 \Delta_G^e(X)}{2|E|} \right)^{\frac{1}{|E|-1}}.$$

I can be shown via straight forward calculations that the update probability can be expressed as:

$$\phi_e(X_{\sim e}) = \mathbb{E}_{X \sim \mu} [X_e | X_{\sim e}] = \frac{\exp(2\beta_0 + \sum_{l=1}^K 2\beta_l |E_l| r_{G_l}(X, e)^{|E_l|-1})}{1 + \exp(2\beta_0 + \sum_{l=1}^K 2\beta_l |E_l| r_{G_l}(X, e)^{|E_l|-1})}.$$

Note that, by definition of $r_G(X, e)$, adding the edge e to $X_{\sim e}$ increases $n^2 N_l(X)$, the scaled homomorphism density of G_l in X , by $2|E_l|r_{G_l}(X, e)^{|E_l|-1}$. The expression for $\phi_e(X_{\sim e})$ is obtained by writing down $\mathcal{H}_\beta(X_e^+) - \mathcal{H}_\beta(X_e^-)$ in terms of $r_{G_l}(X, e)$.

It follows from the definitions that if $X \sim G(n, p)$, then $r_G(X, e) \approx p$ with high probability. If, conversely, it were the case that $r_G(X, e) = p$ for all G_1, \dots, G_K , then the update probability takes the following simpler form (with some abuse of notation) given by $\phi_\beta : [0, 1] \rightarrow [0, 1]$ with

$$(8) \quad \phi_\beta(p) = \frac{\exp(2\beta_0 + \sum_{l=1}^K 2\beta_l |E_l| p^{|E_l|-1})}{1 + \exp(2\beta_0 + \sum_{l=1}^K 2\beta_l |E_l| p^{|E_l|-1})}.$$

Let M_β denote the set of global maximizers of L_β . Let $U_\beta \subset M_\beta$ be the global maximizers where the second derivative of L_β is nonzero. It can be shown that $p^* \in U_\beta$ only if $p^* = \phi_\beta(p^*)$ and $\phi'_\beta(p^*) < 1$. What this implies is that p^* is a *stable fixed point*: if the chain is started at X where $r_G(X, e) \approx p^*$ for all $G \in G_1, \dots, G_K$, then that continues to hold for exponentially many steps. This was shown in Lemma 17 from [1] and is stated in our paper as Lemma 4.9.

We will also need to consider the Glauber dynamics for distributions π that assign zero probability to some graphs.

DEFINITION 2.1 (Glauber Dynamics). Given any probability distribution π over $\mathcal{X} = \{0, 1\}^{\binom{n}{2}}$, we define the Glauber dynamics with respect to π to be the discrete time Markov chain over \mathcal{X} with single step transitions (to obtain X') as follows:

1. Pick a coordinate $E \in \binom{\mathcal{X}}{2}$ uniformly at random.
2. Given the coordinate E , define $X^{\oplus E}$ to be X with edge E flipped, and let

$$(9) \quad X' = \begin{cases} X^{\oplus E} & \text{with probability } \frac{\pi(X^{\oplus E})}{\pi(X^{\oplus E}) + \pi(X)} \quad \text{if } \pi(X) \neq 0 \\ X^{\oplus E} & \text{with probability } 1 \quad \text{if } \pi(X) = 0 \\ X & \text{with probability } \frac{\pi(X)}{\pi(X^{\oplus E}) + \pi(X)} \end{cases}$$

The Glauber dynamics with respect to π is reversible for π ([20]), so in particular π is stationary. When $\pi(X) > 0$ for every X , this reduces to the definition given in Definition 1.1

3. Main Results. Fix $\beta \in \mathbb{R} \times (\mathbb{R}^+)^K$, and recall that M_β denotes the set of global maximizers of L_β . Let $U_\beta \subset M_\beta$ be the global maximizers where the second derivative of L_β is nonzero. Throughout, we will always take $p^* \in U_\beta$. When $|U_\beta| = |M_\beta| = 1$, our main result, stated below in Theorem 3.1, shows that the Glauber dynamics for ERGM when initialized at the $G(n, p^*)$ distribution rapidly approximately mixes as long as the target total variation distance $\delta_0 \geq \exp(-c_\beta n)$ (see Definition 1.4). Note that $|U_\beta| = |M_\beta| = 1$ even in the low temperature regime for Lebesgue almost all β .

In the case that $|U_\beta| > 1$, we show that Glauber dynamics with the same initialization as above can efficiently and approximately sample from the ERGM conditioned on being close in cut metric to the constant p^* graphon. Note that Theorem 1.3 shows that with a very large probability, a sample from the ERGM is close to the constant p graphon for some $p \in U_\beta$. If the probability of being close in cut-metric to each $p \in U_\beta$ under the measure μ is known, then we can initialize the Glauber dynamics to the correct mixture of $(G(n, p))_{p \in U_\beta}$ and show that it mixes rapidly as long as the target total variation distance $\delta_0 \geq \exp(-c_\beta n)$. In this work, we do not consider the problem of estimating these mixture probabilities.

For $\eta > 0$ denote the η -ball in cut metric around p^* by

$$B_\eta^\square(p^*) := \{X \in \mathcal{X} : \delta_\square(X, p^* \mathbf{1}) \leq \eta\}.$$

Let P_η be the kernel of the Glauber dynamics with respect to the measure $\mu(\cdot | B_\eta^\square(p^*))$ and let P be the kernel of the Glauber dynamics with respect to the measure μ .

THEOREM 3.1. *Let $\pi_0 := G(n, p^*)$ for any $p^* \in U_\beta$. Let $\bar{X}_0 \sim \pi_0, \bar{X}_1, \bar{X}_2, \dots$, evolve according to P_η and $X_0 \sim \pi_0, X_1, X_2, \dots$ evolve according to P . There exist positive constants $\eta_0(\beta)$, $c_{\beta,\eta}$, $C_{\beta,\eta}$ and $n_0(\beta, \eta)$ such that whenever $\eta < \eta_0(\beta)$ and $n > n_0(\beta, \eta)$, the following hold:*

1. *(X_0, X_1, \dots, X_T) can be coupled with $(\bar{X}_0, \dots, \bar{X}_T)$ such that with probability at least $1 - TC(\beta, \eta) \exp(-c_{\beta,\eta}n)$, we have*

$$(X_0, X_1, \dots, X_T) = (\bar{X}_0, \bar{X}_1, \dots, \bar{X}_T).$$

2. *Whenever $\delta \geq \exp(-c_{\beta,\eta}n)$, P_η is $(\pi_0, \mu(\cdot | B_\eta^\square(p^*)), C_{\beta,\eta}n^2 \log(n^2/\delta), \delta)$ -mixing.*
3. *If $|U_\beta| = |M_\beta| = 1$, then whenever $\delta \geq \exp(-c_{\beta,\eta}n)$, P is $(\pi_0, \mu, C_{\beta,\eta}n^2 \log(n^2/\delta), \delta)$ -mixing.*

This theorem shows that we can achieve metastable mixing by disregarding a portion of the state space of probability $\exp(-c_{\beta,\eta}n)$ under the measure $\mu(\cdot | B_\eta^\square(p^*))$. One might wonder if this is necessary, and in particular whether it is possible to improve the second item due to $B_\eta^\square(p^*)$ being possibly well-connected. We next answer this question in the negative and gain insight into the structure of the ERGM measure at low temperature.

The paper [1] constructs metastable states where the graph is close to $G(n, p)$ for some p which is a local maximizer of L_β , from which any local Markov chain takes $\exp(\Omega(n))$ time to escape. The large deviations theory based results established in [6] show that when p is not the global minimizer of L_β , then these metastable states collectively have mass $\exp(-\Omega(n^2))$. One might hypothesize that the metastable states can be fully characterized by the behavior of local maximizers of L_β and the cut-metric neighborhoods, and moreover that they have total mass $\exp(-\Omega(n^2))$.

Perhaps surprisingly, it turns out instead that the low-temperature ERGM landscape is remarkably intricate even within the neighborhood $B_\eta^\square(p^*)$ around the global optimizer p^* . We show by construction that this set can contain multiple metastable states which collectively have mass $\exp(-\Theta(n))$ and from which the Glauber dynamics takes $\exp(\Omega(n))$ time to escape. These states are close in cut-metric to the constant graphon p^* and it follows that cut-metric based large deviations analysis cannot capture the intricacies of Markov chain mixing in the ERGM at low-temperatures.

EXAMPLE 3.2. Suppose $K = 1$ and let G_1 be the triangle graph (i.e, the 3 clique). Let $\sigma(x) := e^x / (1 + e^x)$. There exist parameters $\beta_0, \beta_1 \in \mathbb{R} \times \mathbb{R}^+$ and real numbers $p_1^* \neq p_2^*$ such that:

1. p_1^* and p_2^* satisfy $p_i^* = \sigma(2\beta_0 + 6\beta_1(p_i^*)^2)$, $U_\beta = \{p_1^*\}$, and p_2^* is a local maximizer of L_β ;
2. There exists $q^* \in [0, 1]$, $q^* \notin \{p_1^*, p_2^*\}$, such that $q^* = \sigma(2\beta_0 + 6\beta_1 q^* p_1^*)$;
3. Taking $f(x) = \sigma(2\beta_0 + 6\beta_1 x^2)$ and $g(x) = \sigma(2\beta_0 + 6\beta_1 x p_1^*)$, we have $f'(p_1^*) < 1$ and $g'(q^*) < 1$.

We numerically check that the choice $\beta_0 = -1.8$ and $\beta_1 = 2$ has p_1^*, p_2^* , and q^* satisfying the relations above. As shown next, this turns out to imply metastability.

THEOREM 3.3. *Consider Example 3.2 given above. Let $\eta > 0$ be any small enough constant. Let the initial state X_0 be such that $(X_0)_{1j} \sim \text{Ber}(q^*)$ and $(X_0)_{ij} \sim \text{Ber}(p_1^*)$ for $i, j \neq 1$ and $i < j$ are independently distributed. Suppose X_0, X_1, \dots is the trajectory of*

the Glauber dynamics with respect to μ with β as given in Example 3.2. Define the set of graphs $\mathcal{X}_{q,p}(\eta)$ for $q,p \in [0,1]$ by

$$\mathcal{X}_{q,p}(\eta) := \{X \in \mathcal{X} : \delta_{\square}(X, p) \leq \eta/2 \text{ and } |p_1(X) - q| \leq \eta\}.$$

Then the following hold:

1. The set $\mathcal{X}_{q^*,p_1^*}(\eta)$ is metastable: There is a constant $\alpha > 0$ such that

$$\mathbb{P}(\cap_{t \leq \exp(\alpha n)} \{X_t \in \mathcal{X}_{q^*,p_1^*}(\eta)\}) \geq 1 - \exp(-\Omega_\eta(n)).$$

2. The set $\mathcal{X}_{q^*,p_1^*}(\eta)$ has sizable probability:

$$\mu(\mathcal{X}_{q^*,p_1^*}(\eta)) = \exp(-\Theta_\eta(n)).$$

3. Most of the mass lies in $\mathcal{X}_{p_1^*,p_1^*}(\eta)$:

$$\mu(\mathcal{X}_{p_1^*,p_1^*}(\eta)) \geq 1 - \exp(-\Omega_\eta(n)).$$

The theorem is proved in Section 9. Note that the existence of p_2^* is necessary for the model to be in the low-temperature regime and for $q^* \neq p_1^*$ to exist in the result above (else Glauber dynamics would mix in time $O(n^2 \log n)$ as shown by [1]). However, the theorem statement itself does not refer to p_2^* .

In order to see why the set of states described in the theorem above are metastable, consider the first step of Glauber dynamics taking X_0 to X_1 . The number of triangles formed by including an edge $e = (1, j)$ is approximately $nq^*p_1^*$ (which is $6np_1^*q^*$ after counting re-labelings), i.e., $\Delta_{G_1}^e(X_0) \approx \frac{6p_1^*q^*}{n^2}$. Thus, the Glauber dynamics updates this coordinate to 1 with probability $\approx \sigma(2\beta_0 + 6\beta_1 q^* p_1^*) = q^*$. Similarly, if an edge $e = (i, j)$ is to be updated with $i, j \neq 1$, then the number of triangles formed is $n(p_1^*)^2$ (which is $6np_1^*q^*$ after counting re-labelings) i.e., $\Delta_{G_1}^e(X_0) \approx \frac{6(p_1^*)^2}{n^2}$, and the probability of setting this coordinate to 1 is $\approx \sigma(2\beta_0 + 6\beta_1(p_1^*)^2) = p_1^*$. Therefore, the Glauber dynamics update still makes X_1 look approximately like the initial distribution. Not only that, but this is a stable fixed point, which follows from the conditions $f'(p_1^*) < 1$ and $g'(q^*) < 1$.

4. Showing Metastable Mixing for Glauber Dynamics .

4.1. *Couplings, Contraction, and Mixing.* Consider a Markov chain over the finite state space \mathcal{X} and with transition kernel P . Let $d : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}^+$ be such that $\sup_{x,y \in \mathcal{X}} d(x, y) \leq d_{\max}$. We will use the following lemma to establish metastable mixing, proved in Appendix C.

LEMMA 4.1. *Let $A \subseteq \mathcal{X} \times \mathcal{X}$ be such that for $(x, y) \in A$ there exists a γ -contractive coupling Q_{xy} of $P(x, \cdot)$ and $P(y, \cdot)$, i.e. for $(X', Y') \sim Q_{xy}$ we have*

$$\mathbb{E}d(X', Y') \leq (1 - \gamma)d(x, y).$$

Then, given any jointly distributed $(X_0, Y_0) \in \mathcal{X} \times \mathcal{X}$, there exists a coupling between the trajectories $(X_k)_{k \geq 0}$ and $(Y_k)_{k \geq 0}$ of the Markov chain P such that

$$\mathbb{E}d(X_{k+1}, Y_{k+1}) \leq (1 - \gamma)\mathbb{E}d(X_k, Y_k) + d_{\max}p_k,$$

where $p_k := \mathbb{P}((X_k, Y_k) \in A^c) = 1 - \mathbb{P}((X_k, Y_k) \in A)$. Unrolling this recursion, we conclude that

$$\mathbb{E}d(X_K, Y_K) \leq d_{\max} \left[(1 - \gamma)^K + \frac{\sup_{k \leq K} p_k}{\gamma} \right].$$

The following corollary is immediate from the coupling characterization of total variation.

COROLLARY 4.2. In the setting of Lemma 4.1, if additionally $\inf_{x \neq y} d(x, y) \geq d_{\min}$, then $d_{\text{TV}}(X_K, Y_K) \leq \frac{d_{\max}}{d_{\min}} [(1 - \gamma)^K + \gamma^{-1} \sup_{k \leq K} p_k]$.

In essence, the result above shows that whenever two trajectories can be coupled such that with high probability they lie in a set A where a contractive coupling exists, then the laws of their iterates converge until a certain lower threshold. In particular, taking Y_0 to be drawn from the stationary distribution of P , we can establish metastable mixing for X_0, X_1, \dots .

We will use the monotone coupling, defined next.

DEFINITION 4.3 (Monotone coupling). When P is the kernel of the Glauber dynamics with respect to μ , the following coupling between $P(x, \cdot)$ and $P(y, \cdot)$ is called the **monotone coupling**. For any two $x, y \in \mathcal{X}$, we obtain the one step Glauber dynamics updates X', Y' as follows:

1. Pick the update edge $I \in \binom{[n]}{2}$ uniformly at random to be the same for both X' and Y' .
2. Draw $U \sim \text{Unif}([0, 1])$ independent of everything else and set

$$(10) \quad X' = \begin{cases} x^{+I} & \text{if } U \in [0, \phi_I(x_{\sim I})) \\ x^{-I} & \text{otherwise} \end{cases} \quad \text{and} \quad Y' = \begin{cases} y^{+I} & \text{if } U \in [0, \phi_I(y_{\sim I})) \\ y^{-I} & \text{otherwise.} \end{cases}$$

For any two graphs $X, Y \in \mathcal{X}$, the relation $X \preceq Y$ denotes that $X_e \leq Y_e$ for every $e \in \binom{[n]}{2}$. It follows immediately from the definition of the monotone coupling that given $X \preceq Y$, if X', Y' are obtained via the monotone coupling, then $X' \preceq Y'$ almost surely. We next identify a region of the state space over which the coupling is contractive (as required by Lemma 4.1).

4.2. Control of Subgraph Counts Implies Contraction. We will now follow the results established in [1] to show the path coupling of Glauber dynamics and use the notations they introduced. Recall that the update probability under the Glauber dynamics for μ is given by

$$\phi_e(X_{\sim e}) = \frac{\exp(\beta_0 + \sum_{l=1}^K 2\beta_l |E_l| r_{G_l}(X, e)^{|E_l|-1})}{1 + \exp(\beta_0 + \sum_{l=1}^K 2\beta_l |E_l| r_{G_l}(X, e)^{|E_l|-1})},$$

where

$$r_G(X, e) := \left(\frac{n^2 \Delta_G^e(X)}{2|E(G)|} \right)^{\frac{1}{|E(G)|-1}}$$

and $\Delta_G^e(X) = N_G(X^{+e}) - N_G(X^{-e})$.

Let \mathbb{G}_L denote the set of finite simple graphs with at most L vertices (omitting the graph with 1 edge and 2 vertices), where L is a fixed constant satisfying $L > \max_{i \leq K} |V_i|$. Define the set

$$(11) \quad \Gamma_{p^*, \epsilon} := \{X : r_G(X, e) \in [p^* - \epsilon, p^* + \epsilon] \text{ for all } e \in \binom{[n]}{2} \text{ and } G \in \mathbb{G}_L\}.$$

Note that whenever $X \in \Gamma_{p^*, \epsilon}$, $\phi_e(X_{\sim e}) \approx \phi_\beta(p^*) \approx p^*$. That is, each edge updates approximately like $G(n, p^*)$. The significance of $\Gamma_{p^*, \epsilon}$ is that in this set the monotone coupling is contractive, as shown in [1, Lemma 18] and stated next. We will additionally state a theorem in the next subsection that $\Gamma_{p^*, \epsilon}$ has high probability under $\mu(\cdot | B_\eta^\square(p^*))$.

LEMMA 4.4 (Contraction within $\Gamma_{p^*,\epsilon}$, [1]). *Let $p^* \in U_\beta$, $\epsilon > 0$ small enough as a function of β , and n large enough as a function of β, ϵ . Let $\mathcal{A} := \{(x, y) \in \Gamma_{p^*,\epsilon} \times \Gamma_{p^*,\epsilon} : x \preceq y\}$. Let X', Y' be obtained from x, y via one step of the Glauber dynamics under the monotone coupling. There is a constant $c(\beta, \epsilon) > 0$ such that if $(x, y) \in \mathcal{A}$, then*

$$\mathbb{E}d_H(X', Y') \leq \left(1 - \frac{c(\beta, \epsilon)}{n^2}\right)d_H(x, y),$$

and moreover, $X' \preceq Y'$ almost surely.

4.3. *Key Theorem.* We first recall Theorem 1.3 which state that $\cup_{p^* \in U_\beta} B_\eta^\square(p^*)$ has probability $1 - \exp(-\Omega(n^2))$ under the measure μ , that is, most of the mass of μ is concentrated in the cut-metric balls $B_\eta^\square(p^*)$. The following theorem shows that $\mu(\cdot | B_\eta^\square(p^*))$ concentrates over the set $\Gamma_{p^*,\epsilon}$, where path coupling is possible (as per Lemma 4.4).

THEOREM 4.5. *Suppose $p^* \in U_\beta$. Given any $\epsilon > 0$, and any $\eta < c(\beta, \epsilon)$, we must have:*

$$\mu(\Gamma_{p^*,\epsilon} | B_\eta^\square(p^*)) \geq 1 - C(\eta, \epsilon, \beta) \exp(-\Omega_{\beta, \epsilon, \eta}(n)).$$

We prove the theorem in Section 6, modulo lemmas proved via the cavity method in Section 8.

REMARK 4.6. Notice that $r_G(X, e) \in [p^* - \epsilon, p^* + \epsilon]$ uniformly for every e is not implied by $\delta_\square(\tilde{X}, p^*) < \eta$ (for any constant $\eta > 0$). An example is given in Theorem 3.3, where metastability occurs despite being close to the p^* graphon with high probability: The edges emanating from a single vertex prevent uniform concentration of $r_G(X, e)$ in the set $[p^* - \epsilon, p^* + \epsilon]$, but the single vertex neighborhood has a vanishingly small impact on $\delta_\square(\tilde{X}, p^*)$.

While it can easily be proved directly, the following is also a corollary of the above theorem. This is a corollary since any dense Erdős-Rényi random model $G(n, p)$ can be written as an instance of the exponential random graph model. This has p as the single local maximizer for $L_\beta(\cdot)$.

COROLLARY 4.7. Fix any $\epsilon > 0$. Then there exists $\delta_0(\epsilon) > 0$ such that for all $0 < \delta < \delta_0(\epsilon)$, if $Z \sim G(n, p^* + \delta)$, then $\mathbb{P}(Z \in \Gamma_{p^*,\epsilon}) \geq 1 - \exp(-\Omega_\epsilon(n))$.

Even though Theorem 3.2 in [6] considers the probability over the entire space of graphons, we can easily adapt its proof to show the following lemma which considers only the neighborhood $B_{\eta+\delta}^\square(p^*)$ for $p^* \in U_\beta$. A brief sketch of the modifications required is given in Appendix A.

LEMMA 4.8. *Suppose $p^* \in U_\beta$. Then, there exists a constant $c_\beta > 0$ such that whenever $\eta, \delta \in (0, c_\beta)$ are fixed constants independent of n , we have*

$$\frac{\mu(\eta \leq \delta_\square(X, p^*) \leq \eta + \delta)}{\mu(\delta_\square(X, p^*) \leq \eta)} \leq C(\eta, \delta) \exp(-c(\eta, \delta)n^2).$$

4.4. *Metastability.* We intend to invoke Lemma 4.1 to show approximate mixing and prove Theorem 3.1. The prior subsection shows that $G(n, p^*)$ and at $\mu(\cdot | B_\eta^\square(p^*))$ are both within the set $\Gamma_{p^*,\epsilon}$ with high probability. We show that the Glauber chains with these initializations do not leave $\Gamma_{p^*,\epsilon}$ with probability $1 - \exp(-\Omega(n))$ until time $\exp(\Omega(n))$. Some intuition behind this was given in Section 2.3.

We next state Lemma 17 from [1], after adapting it to our situation.

LEMMA 4.9 (Staying in $\Gamma_{p^*, \epsilon}$). *Let $\epsilon > 0$ be a small enough constant independent of n and suppose $p^* \in U_\beta$. Let X_0, X_1, \dots evolve according Glauber dynamics with respect to the measure μ . If $X_0 \in \Gamma_{p^*, \epsilon}$, then for some $\alpha = \alpha(\beta, \epsilon, L)$, we have*

$$\mathbb{P}(X_t \in \Gamma_{p^*, 2\epsilon} \text{ for all } t \leq e^{\alpha n}) \geq 1 - \exp(-\Omega_{\beta, L, \epsilon}(n)).$$

The proofs of the following lemmas are given in Appendix C.

LEMMA 4.10 ($G(n, p^* \pm \epsilon)$ sandwich). *Let $p^* \in U_\beta$. Let constants $\epsilon, \eta > 0$ be such that $\epsilon < \epsilon_0(\beta)$, $\eta < \eta_0(\beta, \epsilon)$, and $n > n_0(\beta, \epsilon, \eta)$. Let $X \sim \mu(\cdot | B_\eta^\square(p^*))$, $\bar{Y} \sim G(n, p^* + \epsilon)$, and $\underline{Y} \sim G(n, p^* - \epsilon)$. Then, there exists a coupling between X, \bar{Y} , and \underline{Y} such that with probability at least $1 - \exp(-\Omega_{\beta, \eta, \epsilon}(n))$*

$$\underline{Y} \preceq X \preceq \bar{Y}.$$

LEMMA 4.11 (Staying in $B_{\eta/2}^\square(p^*)$). *Suppose $p^* \in U_\beta$, $\eta > 0$ such that $\eta < \eta_0(\beta)$ and $n > n_0(\eta, \beta)$. Let $X_0 \sim G(n, p^*)$ and generate the trajectory X_0, \dots, X_T via Glauber dynamics with respect to μ . The entire trajectory X_0, \dots, X_T stays within the ball $B_{\eta/2}^\square(p^*)$ with probability at least $1 - TC(\beta, \eta) \exp(-c(\beta, \eta)n)$.*

5. Proof of Main Result, Theorem 3.1. We now show how the main theorem follows from the various lemmas stated in the last section. Recall that $\pi_0 := G(n, p^*)$, $\bar{X}_0 \sim \pi_0, \bar{X}_1, \bar{X}_2, \dots$ is a trajectory of the Markov chain P_η and $X_0 \sim \pi_0, X_1, X_2, \dots$ is a trajectory of the Markov chain P .

5.1. Proof of Theorem 3.1, Part 1. We will couple the trajectories $\bar{X}_0, \dots, \bar{X}_T$ and X_0, \dots, X_T such that the event $\mathcal{E} := \{(\bar{X}_0, \bar{X}_1, \dots, \bar{X}_T) \neq (X_0, \dots, X_T)\}$ satisfies $\mathcal{E} \subseteq \cup_{t=0}^T \{X_t \in (B_{\eta/2}^\square(p^*))^C\}$. We can then conclude the result from Lemma 4.11. The main observation is that whenever n is large enough as a function of η , if $X \in B_{\eta/2}^\square(p^*)$, then $P_\eta(X, \cdot) = P(X, \cdot)$. We construct the following coupling:

1. $X_0 = \bar{X}_0$ almost surely.
2. X_{t+1}, \bar{X}_{t+1} are drawn from the TV optimal coupling between $P(X_t, \cdot)$ and $P_\eta(\bar{X}_t, \cdot)$.

It is clear that $\{X_{t+1} \neq \bar{X}_{t+1}\} \subseteq \{X_t \neq \bar{X}_t\} \cup \{X_t \in (B_{\eta/2}^\square(p^*))^C\}$. Now, noting that $\{X_0 \neq \bar{X}_0\}$ is the empty event, we conclude that $\{X_1 \neq \bar{X}_1\} \subseteq \{X_0 \in (B_{\eta/2}^\square(p^*))^C\}$. An induction argument with the same basic step shows that $\mathcal{E} \subseteq \cup_{t=0}^T \{X_t \in (B_{\eta/2}^\square(p^*))^C\}$. \square

5.2. Proof of Theorem 3.1, Part 2. Let $\bar{Y}_0 \sim \mu(\cdot | B_\eta^\square(p^*))$ and consider the trajectory $\bar{Y}_0, \bar{Y}_1, \dots, \bar{Y}_T$ with respect to the transition kernel P_η . Note that this is a stationary process since the stationary distribution of P_η is $\mu(\cdot | B_\eta^\square(p^*))$. Similarly, let $Y_0 \sim \mu(\cdot | B_\eta^\square(p^*))$, but with the trajectory Y_1, \dots, Y_T generated with respect to the transition kernel P . Using Lemma 4.8 to bound $\mathbb{P}(\bar{Y}_t \in B_{\frac{\eta}{2}}^\square(p^*))$, a similar proof as in Item 1 shows that

$$(12) \quad d_{\text{TV}}((\bar{Y}_0, \dots, \bar{Y}_T), (Y_0, \dots, Y_T)) \leq T \exp(-\Omega_{\beta, \eta}(n^2)).$$

From Item 1, we have

$$(13) \quad d_{\text{TV}}((\bar{X}_0, \dots, \bar{X}_T), (X_0, \dots, X_T)) \leq T \exp(-\Omega_{\beta, \eta}(n)).$$

These last two displays allow us to consider the total variation distance between the distributions of X_T and Y_T instead of \bar{X}_T and \bar{Y}_T . Let $\epsilon > 0$ be small enough to satisfy the

conditions in Lemma 4.9. By Lemma 4.10 and Theorem 4.5, we conclude that for some $0 < \delta < \delta_0(\beta, \epsilon, \eta)$, $\eta < \eta_0(\delta, \epsilon, \beta)$, and $n > n_0(\delta, \beta, \eta, \epsilon)$, we can couple $Z_0 \sim G(n, p^* + \delta)$ with X_0 and Y_0 such that with probability at least $1 - \exp(-\Omega_{\beta, \delta, \eta, \epsilon}(n))$, the following hold:

1. $X_0 \preceq Z_0$ and $Y_0 \preceq Z_0$, and
2. $X_0, Y_0, Z_0 \in \Gamma_{p^*, \epsilon}$.

Now, we consider the Markov chain Z_0, Z_1, \dots, Z_T with respect to P . We consider the monotone coupling between Y_0, \dots, Y_T and Z_0, \dots, Z_T as in Definition 4.3: conditional on $Y_0 \preceq Z_0$, we have $Y_t \preceq Z_t$ almost surely for every $t \leq T$. Recall $\Gamma_{p^*, \epsilon}$ from (11) and the set $\mathcal{A} := \{(x, y) \in \Gamma_{p^*, \epsilon} \times \Gamma_{p^*, \epsilon} : x \preceq y\}$ defined in Lemma 4.4. Now, $(Y_t, Z_t) \in \mathcal{A}^c$ implies either $Y_0 \not\preceq Z_0$, $Y_t \notin \Gamma_{p^*, \epsilon}$, or $Z_t \notin \Gamma_{p^*, \epsilon}$. The first of these is ruled out by monotonicity of the coupling. For the latter two, Lemma 4.9 shows that starting in $\Gamma_{p^*, \epsilon}$ the trajectory stays there for some time, and it follows that

$$\mathbb{P}((Y_t, Z_t) \in \mathcal{A}^c) \leq \exp(-\Omega_{\beta, \delta, \eta, \epsilon}(n)).$$

Applying Lemmas 4.1 and 4.4 with Hamming distance d_H over \mathcal{X} , we conclude that whenever $T \leq \exp(c_0 n)$ for small enough c_0 as a function of $\epsilon, \delta, \eta, \beta$

$$\mathbb{E}d(Y_T, Z_T) \leq \frac{n^2}{2} \left[\left(1 - \frac{c(\epsilon, \beta)}{n^2}\right)^T + \frac{n^2}{c(\epsilon, \beta)} \exp(-\Omega_{\eta, \beta, \epsilon, \delta}(n)) \right].$$

Similarly, we have

$$\mathbb{E}d(X_T, Z_T) \leq \frac{n^2}{2} \left[\left(1 - \frac{c(\epsilon, \beta)}{n^2}\right)^T + \frac{n^2}{c(\epsilon, \beta)} \exp(-\Omega_{\eta, \beta, \epsilon, \delta}(n)) \right].$$

Combining the two displays above, the coupling characterization of TV distance implies

$$\begin{aligned} d_{\text{TV}}(X_T, Y_T) &\leq \mathbb{P}(X_T \neq Y_T) \leq \mathbb{P}(X_T \neq Z_T) + \mathbb{P}(Z_T \neq Y_T) \\ &\leq \mathbb{P}(d(X_T, Z_T) > 1) + \mathbb{P}(d(Y_T, Z_T) > 1) \\ &\leq \mathbb{E}d(X_T, Z_T) + \mathbb{E}d(Y_T, Z_T) \\ (14) \quad &\leq n^2 \left[\left(1 - \frac{c(\epsilon, \beta)}{n^2}\right)^T + \frac{n^2}{c(\epsilon, \beta)} \exp(-\Omega_{\eta, \beta, \epsilon, \delta}(n)) \right]. \end{aligned}$$

Now, we will allow ϵ, δ to be small enough constants as a function of β such that Lemmas 4.9 and 4.4 hold. Whenever η is small enough as a function β and n is large enough, combining Equations (12), (13), and (14), yields

$$d_{\text{TV}}(\bar{X}_T, \bar{Y}_T) \leq n^2 \left[\left(1 - \frac{c(\epsilon, \beta)}{n^2}\right)^T + \frac{n^2 T}{c(\epsilon, \beta)} \exp(-\Omega_{\eta, \beta}(n)) \right].$$

This yields the second part of the theorem statement by considering $T = C_{\beta, \eta} n^2 \log(n^2/\delta)$. \square

5.3. Proof of Theorem 3.1, Part 3. Whenever $|U_\beta| = 1$, the concentration result of [6] (stated here as Theorem 1.3) implies that

$$d_{\text{TV}}(\mu, \mu(\cdot | B_\eta^\square(p^*))) \leq \exp(-\Omega_{\beta, \eta}(n^2)).$$

Meta-stable mixing to μ follows from the second part and the triangle inequality. \square

6. Uniform Subgraph Concentration and Proof of Theorem 4.5. In this section we reduce the proof of Theorem 4.5 on the concentration of $\Delta_G^e(X) := N_G(X^{+e}) - N_G(X^{-e})$ to control of both the vertex degrees $p_u(X)$ and common neighbors $p_{uv}(X)$. The latter are stated here as corollaries and will be proved via the cavity method in Section 8.

Theorem 4.5 states that if we sample $X \sim \mu(\cdot | B_\eta^\square(p^*))$, then $X \in \Gamma_{p^*, \epsilon}$ with probability $1 - C(\beta, \epsilon, \eta) \exp(-\Omega_{\beta, \epsilon, \eta}(n))$. Unpacking the definitions, it suffices to show that for some small enough $h(\epsilon, L)$,

$$(15) \quad \left| \Delta_G^e(X) - \frac{2|E(G)|(p^*)^{|E(G)|-1}}{n^2} \right| \leq \frac{h(\epsilon, L)}{n^2}$$

for every $G \in \mathbb{G}_L$ and $e \in \binom{[n]}{2}$, with probability at least $1 - C(\beta, \eta, \epsilon, L) \exp(-\Omega_{\beta, \eta, \epsilon, L}(n))$.

We start with a lemma (proved in Appendix B) which shows that (15) (and hence Theorem 4.5) follows from uniform control of both the vertex degrees $p_u(X)$ and common neighbors $p_{uv}(X)$. Some notation is needed. Given a fixed graph $G = ([k], E)$ and vertices $i, j \in [k]$ such that $(i, j) \in E$, let

$$(16) \quad E_{ij}(G) := \{l : l \in [k], l \neq i, l \neq j, (l, i) \in E, (l, j) \in E\} \quad \text{and} \quad d_{ij}(G) = |E_{ij}(G)|.$$

In words, $d_{ij}(G)$ is the number of common neighbors of vertices i and j in G .

LEMMA 6.1. *Suppose $X \in \mathcal{X}$, $e = (u, v)$, and $p^* \in [0, 1]$ are such that $\sup_{u \in [n]} |p_u(X) - p^*| \leq \epsilon$. For any fixed graph G and for some constant C_G depending only on G , we have*

$$\left| \Delta_G^e(X) - \frac{2}{n^2} \sum_{(i, j) \in E(G)} \left(\frac{p_{uv}(X)}{(p^*)^2} \right)^{d_{ij}} (p^*)^{|E(G)|-1} \right| \leq \frac{C_G}{n^2} \left(\epsilon + \delta_\square(\tilde{X}, p^*) + n^{-1} \right).$$

COROLLARY 6.2. Let $C(L) = \max_{G \in \mathbb{G}_L} C_G$ for C_G in Lemma 6.1 and $h(\epsilon, L)$ be as in (15). Let $a(\epsilon, L) = h(\epsilon, L)/3C(L)$. Then $X \in \Gamma_{p^*, \epsilon}$ holds if

1. X is $a(\epsilon, L)$ -close to the constant graphon p^* in the cut-metric,
2. $p_u(X)$ is uniformly close to p^* for every vertex u , i.e., $\sup_u |p_u(X) - p^*| \leq a(\epsilon, L)$, and
3. $p_{uv}(X)$ is close to $(p^*)^2$ uniformly for every pair of vertices $u \neq v$, i.e.,

$$(17) \quad \sup_{u \neq v} |p_{uv}(X) - (p^*)^2| \leq p^* a(\epsilon, L)/3L.$$

It follows that Theorem 4.5 is proved if these conditions are each shown to hold for $X \sim \mu(\cdot | B_\eta^\square(p^*))$ with probability at least $1 - C(\beta, \epsilon, \eta, L) \exp(-\Omega_{\beta, \epsilon, \eta, L}(n))$ whenever we choose $0 < \eta < c(\beta, \epsilon)$ to be a small enough constant. Item 1 holds by Theorem 1.3. We address Items 2 and 3 below.

6.1. Uniform Control of Degrees. Section 8 develops the cavity method for the ERGM and demonstrates the following uniform control on vertex degrees.

COROLLARY 6.3. Suppose $p^* \in U_\beta$ and let $\epsilon > 0$ be an arbitrary fixed constant. Then, for any $0 < \eta < c(\beta, \epsilon)$ and $n > n_0(\beta, \epsilon, \eta)$, we have:

$$\mu \left(\sup_{u \in [n]} |p_u(X) - p^*| \leq \epsilon \mid B_\eta^\square(p^*) \right) \geq 1 - \exp(-\Omega_{\beta, \epsilon}(n)).$$

We take $\eta < c(\beta, a(\epsilon, L))$ in Corollary 6.3, implying that if we sample $X \sim \mu(\cdot | B_\eta^\square(p^*))$ then Item 2 of Corollary 6.2 holds with probability $1 - C(\beta, \epsilon, \eta, L) \exp(-\Omega_{\beta, \epsilon, \eta, L}(n))$.

It remains to show Item 3 of Corollary 6.2. As a step towards this, it turns out that if $p_u(X)$ is close to p^* uniformly for *every* vertex u (as stated in Corollary 6.3) and X is close to the constant graphon p^* , then *most* of the common neighbor counts $p_{uv}(X)$ are close to $(p^*)^2$. This follows from the definition of cut-metric.

COROLLARY 6.4. Suppose that $p^* \in U_\beta$. Given arbitrary $\epsilon, \delta > 0$, suppose $\eta < c(\beta, \epsilon, \delta)$ and $n > n_0(\delta, \epsilon, \beta, \eta)$. Then, for every $u \in [n]$, there exists a random set $S_u \subseteq [n] \setminus \{u\}$ such that $|S_u| \leq \delta n$ and

$$\mu \left(\sup_{u \in [n]} \sup_{v \in S_u^c} |p_{uv}(X) - (p^*)^2| \leq \epsilon \mid B_\eta^\square(p^*) \right) \geq 1 - \exp(-\Omega_{\beta, \epsilon}(n)).$$

We refer to Appendix D.1 for the proof.

Next, to establish Equation (17), we boost control of $p_{uv}(X)$ from most pairs u, v to all pairs.

6.2. Uniform Control of Common Neighbors. Let X' be obtained from $X \sim \mu(\cdot | B_\eta^\square(p^*))$ via one step of the Glauber dynamics with respect to $\mu(\cdot | B_\eta^\square(p^*))$, so that we also have $X' \sim \mu(\cdot | B_\eta^\square(p^*))$. As shown in Lemma 4.8, with high probability X is $\eta/2$ away from the boundary of $B_\eta^\square(p^*)$ (that is, $X \in B_{\frac{\eta}{2}}^\square(p^*)$) and the expected Glauber update for p_{uv} with respect to $\mu(\cdot | B_\eta^\square(p^*))$ is the same as with respect to $\mu(\cdot)$, which is

$$\begin{aligned} \mathbb{E} [p_{uv}(X') | X] &= \left(1 - \frac{2}{N}\right) p_{uv}(X) \\ &\quad + \frac{1}{N} \left(n^{-1} \sum_{w \in [n] \setminus \{u, v\}} \phi_{uw}(X_{\sim uw}) X_{vw} + \phi_{vw}(X_{\sim vw}) X_{uw} \right). \end{aligned}$$

The next lemma shows that under the conditions shown in Corollaries 6.3 and 6.4 to hold with high probability for $X \sim \mu(\cdot | B_\eta^\square(p^*))$, each of the normalized sums in the last displayed equation is close to $(p^*)^2$ uniformly for every $u \neq v$. The Glauber dynamics with respect to $\mu(\cdot | B_\eta^\square(p^*))$, therefore, tries to regress every $p_{uv}(X)$ close to $(p^*)^2$.

LEMMA 6.5. Suppose $X \in \mathcal{X}$, $p^* \in [0, 1]$ are such that the following conditions hold:

1. $\sup_{u \in [n]} |p_u(X) - p^*| \leq \epsilon$, and
2. For every $u \in [n]$, there exist sets $S_u \subseteq [n]$ such that $|S_u| \leq \delta n$ and

$$\sup_{u \in [n]} \sup_{w \in S_u^c} |p_{uw}(X) - (p^*)^2| \leq \epsilon.$$

Then, for every $u, v \in [n]$, we have

$$\left| n^{-1} \sum_{w \in [n] \setminus \{u, v\}} \phi_{uw}(X_{\sim uw}) X_{vw} - (p^*)^2 \right| \leq C_\beta (\epsilon + \delta + \delta_\square(X, p^*) + n^{-1}).$$

The lemma is proved in the next subsection.

Intuitively, this suggests that the stationary distribution of the Glauber dynamics, $\mu(\cdot | B_\eta^\square(p^*))$, should be such that $p_{uv}(X) \approx (p^*)^2$ for every $u \neq v$ with high probability. The next lemma formalizes this sentiment using Stein's method for concentration developed in [5, Theorem 1.5]. We refer to Appendix B for its proof.

LEMMA 6.6. For any $u, v \in [n]$ such that $u \neq v$, define

$$g_{uv}(X) := p_{uv}(X) - \frac{1}{2n} \sum_{w \in [n] \setminus \{u, v\}} \left(\phi_{uw}(X_{\sim uw}) X_{vw} + \phi_{vw}(X_{\sim vw}) X_{uw} \right).$$

Then, for any $\gamma > 0$ which is independent of n , we have

$$(18) \quad \mu \left(\{|g_{uv}(X)| > \gamma\} \cup (B_{\eta/2}^{\square}(p^*))^{\complement} \middle| B_{\eta}^{\square}(p^*) \right) \leq C(\eta, \beta) \exp(-\Omega_{\gamma, \eta, \beta}(n)).$$

COROLLARY 6.7. Item 3 of Corollary 6.2 holds with the desired probability. \square

PROOF. Combining the last two lemmas with Corollary 6.3 proves (17). \square

6.3. *Proof of Lemma 6.5.* First, note that by definition

$$\phi_{uw}(X_{\sim uw}) = \frac{\exp(\sum_{i=0}^K n^2 \beta_i \Delta_i^{uw}(X))}{1 + \exp(\sum_{i=0}^K n^2 \beta_i \Delta_i^{uw}(X))}.$$

Now, suppose $w \in S_u^{\complement}$. Then, using Lemma 6.1, we conclude that

$$\sum_{i=0}^K n^2 \beta_i \Delta_i^{uw}(X) = \sum_{i=0}^K 2|E_i| \beta_i(p^*)^{|E_i|-1} \pm O_{\beta}(\epsilon + \delta_{\square}(X, p^*) + n^{-1}).$$

Using the fact that the function $x \rightarrow \frac{e^x}{1+e^x}$ is 1-Lipschitz, we have for $w \in S_u^{\complement}$ that

$$\phi_{uw}(X_{\sim uw}) = \frac{\exp(\sum_{i=0}^K 2|E_i| \beta_i(p^*)^{|E_i|-1})}{1 + \exp(\sum_{i=0}^K 2|E_i| \beta_i(p^*)^{|E_i|-1})} \pm O_{\beta}(\epsilon + \delta_{\square}(X, p^*) + n^{-1}).$$

Now, the fact that $p^* \in U_{\beta}$ implies that $L'_{\beta}(p^*) = 0$. It can be easily checked that this implies $p^* = \phi_{\beta}(p^*)$. It follows that whenever $w \in S_u^{\complement}$,

$$\phi_{uw}(X_{\sim uw}) = p^* \pm O_{\beta}(\epsilon + \delta_{\square}(X, p^*) + n^{-1}).$$

An application of the triangle inequality now shows that

$$\left| n^{-1} \sum_{w \in [n] \setminus \{u, v\}} \phi_{uw}(X_{\sim uw}) X_{vw} - p^* p_v(X) \right| \leq \frac{|S_u|}{n} + O_{\beta}(\epsilon + \delta_{\square}(X, p^*) + n^{-1}).$$

The assumption that $|S_u| \leq \delta n$ and the fact that $|p_v(X) - p^*| \leq \epsilon$ imply the result. \square

7. Some Graphon Estimates. We state three technical lemmas below, whose proofs appear in Appendix A.

LEMMA 7.1. For any graph X over n vertices, $p \in [0, 1]$, and $\delta > 0$, there exists a set $S \subseteq [n]$ such that $|S| \leq \delta n$ and

$$(19) \quad \sup_{u \in S^{\complement}} |p_u(X) - p| \leq \frac{2\delta_{\square}(\tilde{X}, p\mathbf{1})}{\delta}.$$

It follows as an application that for $X \sim \text{ERGM}(n, \beta)$ and any given fixed constants $\delta, \eta > 0$, with probability at-least $1 - C(\eta) \exp(-c(\eta)n^2)$, there exists a (random) set $S \subseteq [n]$ such that $|S| \leq \delta n$ and

$$\inf_{p^* \in U_{\beta}} \sup_{u \in S^{\complement}} |p_u(X) - p^*| \leq \frac{2\eta}{\delta}.$$

LEMMA 7.2. *If $Y \preceq X \preceq Z$, then $\delta_{\square}(\tilde{X}, p\mathbf{1}) \leq \max(\delta_{\square}(\tilde{Y}, p\mathbf{1}), \delta_{\square}(\tilde{Z}, p\mathbf{1}))$.*

LEMMA 7.3. *Consider any fixed graph $G = (V, E)$. For any vertex i , let d_i denote the degree of the vertex i . Suppose $p, q \in [0, 1]$. If $G = G_0$, the graph consisting of a single edge, then*

$$(20) \quad \sum_{i=1}^{|V|} q^{d_i} p^{|E|-d_i} = 2q^{|E|} + (|V|-2)p^{|E|}.$$

If G is connected and $G \neq G_0$, then there is a constant C depending only on $(d_i)_{i \in V}$ and $|E|$ such that

$$\begin{aligned} & 2q^{|E|} + (|V|-2)p^{|E|} - C|p-q| \\ & \leq \sum_{i=1}^{|V|} q^{d_i} p^{E-d_i} \leq 2q^{|E|} + (|V|-2)p^{|E|} - \zeta(|p-q|), \end{aligned}$$

where $\zeta : [0, 1] \rightarrow \mathbb{R}^+$ is a continuous function depending only on G such that $\zeta(x) > 0$ for $x \neq 0$.

8. The Cavity Method. In this section we address the degrees and show that every vertex u has nearly the same degree $p_u(X) \approx p^* \in U_{\beta}$ with high probability for $X \sim \mu(\cdot | B_{\eta}^{\square}(p^*))$. While the cut-metric based convergence does not allow us to control all the degrees, it is nevertheless possible to conclude that a large portion of the vertices have degree $p_u(X) \approx p^*$. We boost this to a uniform statement, in Theorem 8.9 and Corollary 6.3, via the *cavity method*: most of the vertices and the corresponding edges are conditioned on being close to the constant graphon $p^* \mathbf{1}$, which generates the *mean field* with which the remaining cavity vertices interact. We can then reason about the behavior of the cavity vertices.

We start by adapting several of the graphon definitions to incorporate a cavity.

8.1. Restricted Homomorphism Densities and Restricted Cut Metric. Recall from Section 2.2 the function representative $f^X(x, y)$ of a graph X over n vertices. We will need the homomorphism density of a graph forced to contain a particular vertex u of X . To that end, for every $u \in [n]$, define the event

$$A_u^{k,i} := \left\{ x \in [0, 1]^k : x_i \in \left[\frac{u-1}{n}, \frac{u}{n} \right) \right\}$$

and

$$(21) \quad A_u^k = \left\{ x \in [0, 1]^k : x_i \in \left[\frac{u-1}{n}, \frac{u}{n} \right) \text{ for some } i \in [k] \right\} = \bigcup_{i \in [k]} A_u^{k,i}.$$

For the sake of clarity in the results below, given a fixed graph G , we take its vertex set $V(G) = [k]$ during calculations.

DEFINITION 8.1 (Homomorphism density w.r.t. a vertex). Define the homomorphism density of G in X with respect to vertex u (which counts only homomorphisms which include the vertex u) as

$$(22) \quad N_G(X; u) := \int_{[0, 1]^k} \mathbb{1}(A_u^k) \prod_{(i, j) \in E(G)} f^X(x_i, x_j) \prod_{i=1}^k dx_i.$$

For ease of computation, we also introduce the quantity

$$(23) \quad N_G^0(X; u) := \sum_{l=1}^k \int_{[0,1]^k} \mathbb{1}(A_u^{k,l}) \prod_{(i,j) \in E(G)} f^X(x_i, x_j) \prod_{i=1}^k dx_i.$$

The next lemma follows from elementary arguments; see Appendix A.

LEMMA 8.2. $0 \leq N_G^0(X, u) - N_G(X, u) \leq |V(G)|^3/n^2$.

We will now show that whenever the graphon corresponding to X is close to a constant graphon $p\mathbf{1}$, then $N_G(X, u)$ can be approximated as a polynomial of $p_u(X)$ and p . This will allow us to control the fine-grained structure of X in terms of the counts $N_G(X, u)$ just based on nearness to a constant graphon and the normalized degrees of the vertices of X . The proof, given in Appendix A, follows from a slight modification of the proof technique of [2, Lemma 4.4], which establishes the continuity of homomorphism densities with respect to the graphon metric via a repeated application of the triangle inequality.

LEMMA 8.3. *Suppose $D_G = (d_1, \dots, d_{|V(G)|})$ is the degree sequence of the fixed graph G considered above. For any graph X with vertex set $[n]$ and $u \in [n]$, we have*

$$\left| N_G(X; u) - n^{-1} \sum_{d \in D_G} p_u(X)^d p^{|E(G)|-d} \right| \leq |V(G)| |E(G)| \frac{\delta_{\square}(\tilde{X}, p)}{n} + \frac{|V(G)|^3}{n^2}.$$

Let $S \subset [n]$ be the “cavity set”. Define

$$A_S^k := \cup_{u \in S} A_u^k \quad \text{and} \quad A_S^{k,i} := \cup_{u \in S} A_u^{k,i}.$$

We have that

$$N_G(X) = N_G(X; S) + \int_{[0,1]^k} \mathbb{1}\{(A_S^k)^{\complement}\} \prod_{(i,j) \in E(G)} f^X(x_i, x_j) \prod_{i=1}^k dx_i,$$

where the subgraph count $N_G(X; S)$ restricts to subgraphs containing a cavity vertex,

$$N_G(X; S) := \int_{[0,1]^k} \mathbb{1}\{A_S^k\} \prod_{(i,j) \in E(G)} f^X(x_i, x_j) \prod_{i=1}^k dx_i.$$

The proofs of the next two lemmas are deferred to Appendix A.

LEMMA 8.4. *Let X be any simple graph with vertex set $[n]$ and let $S \subset [n]$ be arbitrary. Then*

$$\left| N_G(X; S) - \sum_{u \in S} N_G(X; u) \right| \leq \frac{k^3 |S|^2}{n^2}.$$

We now define the graphon metric restricted to S^{\complement} .

DEFINITION 8.5 (Restricted Graphon Metric). *Let $p^* \in [0, 1]$. Define*

$$(24) \quad f^{X, S, p^*}(x_1, x_2) = \begin{cases} f^X(x_1, x_2) & \text{if } \lceil nx_1 \rceil, \lceil nx_2 \rceil \in S^{\complement} \\ p^* & \text{otherwise,} \end{cases}$$

and let \tilde{X}^{S,p^*} be the graphon corresponding to f^{X,S,p^*} . We define the restricted cut metric to be

$$\delta_{\square}^{S,p^*}(\tilde{X}, p^*) := \delta_{\square}(\tilde{X}^{S,p^*}, p^*).$$

The restricted graphon distance to p^* can be approximated in terms of the unrestricted distance:

LEMMA 8.6. *For any $S \subseteq [n]$, we have*

$$\delta_{\square}(\tilde{X}, p^*) - \frac{|S|(2n - |S|)}{n^2} \leq \delta_{\square}^{S,p^*}(\tilde{X}, p^*) \leq \delta_{\square}(\tilde{X}, p^*).$$

We are now ready to establish the cavity decomposition of the Hamiltonian.

8.2. *Cavity Decomposition of the Hamiltonian.* It will be convenient to let $r := |S|$. Given a simple graph X over n vertices, we define $\bar{p}_u(X)$ to be the number of edges from vertex $u \in [n]$ to the set S^c , normalized by n :

$$\bar{p}_u(X) = \frac{1}{n} \sum_{v \in S^c} X_{uv}.$$

Here and throughout we hide the dependence on S to streamline the notation. Additionally, whenever it is clear, we will denote $\bar{p}_u(X)$ by \bar{p}_u . Note that $|\bar{p}_u(X) - p_u(X)| \leq |S|/n$.

Denote the portion of the Hamiltonian associated to the cavity by

$$\mathcal{H}_{\beta}^{\text{cav}}(X; S) := \sum_{i=0}^K n^2 \beta_i N_i(X; S),$$

which is the same as \mathcal{H}_{β} except that the homomorphism densities are restricted to have at least one vertex in the set S . Denote the rest of the Hamiltonian by

$$\mathcal{H}_{\beta}^{\text{mean}}(X; S) = \mathcal{H}_{\beta}(X) - \mathcal{H}_{\beta}^{\text{cav}}(X; S).$$

We next bound the difference between $\mathcal{H}_{\beta}(X)$ and $\mathcal{H}_{\beta}^{\text{mean}}(X; S)$.

LEMMA 8.7 (Cavity Decomposition). *Assume that $\beta_i > 0$ for some $i = 1, \dots, K$. We have the following upper and lower bounds.*

1. *Upper Bound:*

$$\begin{aligned} \mathcal{H}_{\beta}(X) &\leq \mathcal{H}_{\beta}^{\text{mean}}(X; S) + O_{\beta}(nr\delta_{\square}^{S,p^*}(\tilde{X}, p^*) + r^2) + \sum_{i=0}^K \beta_i nr |V_i| (p^*)^{|E_i|} \\ (25) \quad &+ 2n \sum_{u \in S} \left[L_{\beta}(\bar{p}_u) + I(\bar{p}_u) - L_{\beta}(p^*) - I(p^*) - \zeta_{\beta}(|\bar{p}_u - p^*|) \right] \end{aligned}$$

2. *Lower Bound:*

$$\begin{aligned} \mathcal{H}_{\beta}(X) &\geq \mathcal{H}_{\beta}^{\text{mean}}(X; S) - O_{\beta}(nr\delta_{\square}^{S,p^*}(\tilde{X}, p^*) + r^2) + \sum_{i=0}^K \beta_i nr |V_i| (p^*)^{|E_i|} \\ (26) \quad &+ 2n \sum_{u \in S} \left[L_{\beta}(\bar{p}_u) + I(\bar{p}_u) - L_{\beta}(p^*) - I(p^*) - C_{\beta}|\bar{p}_u - p^*| \right] \end{aligned}$$

Here ζ_{β} is a function with the same properties of the function ζ in Lemma 7.3

PROOF. Let D_i be the tuple of the degrees of vertices in G_i (as in Lemma 8.3). In the equations below, we will take $\sum_{d_i \in D_i}$ to mean summation over all elements of the tuple. We have

$$\begin{aligned}
\mathcal{H}_\beta(X) &= \mathcal{H}_\beta^{\text{mean}}(X; S) + \mathcal{H}_\beta^{\text{cav}}(X; S) \\
&= \mathcal{H}_\beta^{\text{mean}}(X; S) + \sum_{u \in S} \sum_{i=0}^K \beta_i n^2 N_i(X; u) \pm O_\beta(r^2) \\
&= \mathcal{H}_\beta^{\text{mean}}(X; S) + \sum_{u \in S} \sum_{i=0}^K \sum_{d_i \in D_i} \beta_i n p_u^{d_i}(p^*)^{E_i - d_i} \pm O_\beta(nr\delta_\square(\tilde{X}, p^*) + r^2) \\
(27) \quad &= \mathcal{H}_\beta^{\text{mean}}(X; S) + \sum_{u \in S} \sum_{i=0}^K \sum_{d_i \in D_i} \beta_i n \bar{p}_u^{d_i}(p^*)^{E_i - d_i} \pm O_\beta(nr\delta_\square(\tilde{X}, p^*) + r^2).
\end{aligned}$$

The first step is by the definition of the cavity Hamiltonian. The second step uses Lemma 8.4 to approximate $N(G; S)$ and the third step uses Lemma 8.3 to approximate $N_i(X; u)$. In the fourth step, we have used the fact that $|\bar{p}_u - p_u| \leq r/n$.

We now apply Lemma 7.3 to the second term of Equation (27), yielding

$$\begin{aligned}
\sum_{i=0}^K \sum_{d_i \in D_i} \beta_i \bar{p}_u^{d_i}(p^*)^{E_i - d_i} &\leq \sum_{i=0}^K \beta_i \left[2\bar{p}_u^{|E_i|} + (|V_i| - 2)(p^*)^{|E_i|} \right] - \zeta_\beta(|\bar{p}_u - p^*|) \\
(28) \quad &= \sum_{i=0}^K \beta_i |V_i|(p^*)^{|E_i|} + 2 \left[L_\beta(\bar{p}_u) + I(\bar{p}_u) - L_\beta(p^*) - I(p^*) - \zeta_\beta(|\bar{p}_u - p^*|) \right].
\end{aligned}$$

In the first step we have used Lemma 7.3 and the fact that for G_0 the inequality is an equality which allows for all $\beta_0 \in \mathbb{R}$. For $i > 0$, notice that $\beta_i \geq 0$ and the inequality goes in the right direction. The function ζ_β is as defined in the statement of Lemma 7.3 and exists since $\beta_i > 0$ for some $i \in [K]$. To see this, observe that the Equation 20 is an equality when $G = G_0$. Therefore, in order to establish the strict inequality involving ζ as shown in Lemma 7.3, we need at-least one of the $\beta_i > 0$.

The upper bound in the lemma statement follows by combining Equations (27) and (28) along with Lemma 8.6 to show that

$$nr\delta_\square(\tilde{X}, p^*) \leq nr\delta_\square^{S, p^*}(\tilde{X}, p^*) + 2r^2.$$

The lower bound on the Hamiltonian follows from a similar argument by replacing the upper bound in Lemma 7.3 with the lower bound. \square

8.3. Controlling Degrees of Cavity Vertices. Given a sequence $\bar{q}_u \in \{0, 1/n, 2/n, \dots, 1\}$ for $u \in S$, we define the tuple $\mathbf{q}_S = (\bar{q}_u)_{u \in S}$. Given arbitrary and fixed $\eta > 0$, $p^* \in [0, 1]$, we define the events

$$(29) \quad A(S, \mathbf{q}_S, p^*, \eta) = \{X : \bar{p}_u(X) = \bar{q}_u \text{ for } u \in S\} \cap \{\delta_\square^{S, p^*}(\tilde{X}, p^*) \leq \eta\}$$

and

$$(30) \quad B(S, p^*, \eta) := \{\delta_\square^{S, p^*}(\tilde{X}, p^*) \leq \eta\}.$$

Note that by definition, $B(\emptyset, p^*, \eta) = B_\eta^\square(p^*)$. We want to show that whenever $p^* \in U_\beta$, if \mathbf{q}_S is not close to p^* , then the event $A(S, \mathbf{q}_S, p^*, \eta)$ has exponentially small probability compared to the event $B(S, p^*, \eta)$, whenever η and S are small enough.

We now note that $\mathcal{H}_\beta^{\text{mean}}(X; S)$ and $\delta_{\square}^{S, p^*}(\tilde{X}, p^*)$ depend only on X_{uv} for $u, v \in S^c$. Therefore, whenever $|S|$ is small, we will think of $\mathcal{H}_\beta^{\text{mean}}(X; S)$ as the mean field which controls the behavior of the cavity, i.e., the edges emanating from the vertices in S . Now, fixing X such that $\delta_{\square}^{S, p^*}(\tilde{X}, p^*) \leq \eta$, we look at the joint law of (X_{uv}) such that at least one of u or v is in the set S . By $X_{\text{mean}}(S)$ we denote the coordinates $(X_{uv})_{u, v \in S^c}$. We denote the rest of the coordinates by $X_{\text{cav}}(S)$. Therefore, we want to understand the conditional law $X_{\text{cav}}(S)|X_{\text{mean}}(S)$ under the measure μ . We first record the following combinatorial lemma, whose proof can be found in Appendix A.

LEMMA 8.8. *Suppose $r/n \leq 1/2$ and $X_{\text{mean}}(S)$ is fixed. Let the count of $X_{\text{cav}}(S)$ such that $\bar{p}_u(X) = q_u$ for $u \in S$ be denoted by $H_{\text{cav}}(\mathbf{q}_S)$. $H_{\text{cav}}(\mathbf{q}_S)$ satisfies*

$$\exp\left(-r^2\left[4 + 2\log\left(\frac{n}{r}\right)\right] - \frac{r}{2}\log(2n)\right) \leq H_{\text{cav}}(\mathbf{q}_S) \exp\left(2n \sum_{u \in S} I(q_u)\right) \leq 1.$$

Below we present the main result of this section.

THEOREM 8.9. *Let $p^* \in U_\beta$ be such that $r/n < p^* < 1 - r/n$ and $r < n/2$. Given any $\mathbf{q}_S = (\bar{q}_u)_{u \in S}$ as defined above for n sufficiently large as a function of β, η , we have*

$$\frac{\mu(A(S, \mathbf{q}_S, p^*, \eta))}{\mu(B(S, p^*, \eta))} \leq \exp\left(2n \sum_{u \in S} \left(L_\beta(\bar{q}_u) - L_\beta(p^*) - \zeta_\beta(|\bar{q}_u - p^*|)\right) + \text{lower order}\right),$$

where lower order = $O_\beta(nr\eta + r^2 \log(n/r) + r \log n)$.

PROOF. Fix a $p^* \in U_\beta$. For the sake of convenience, only in this proof, we will denote $A(S, \mathbf{q}_S, p^*, \eta)$ by A , $B(S, p^*, \eta)$ by B and O_β by O . Let C be the event $\{X : \bar{p}_u(X) = \bar{q}_u \text{ for } u \in S\}$. Let n be large enough so that the sets A, B and C are non-empty. Note that $\mathbb{1}\{X \in B\}$ is a function of X_{mean} and $\mathbb{1}\{X \in C\}$ is a function of X_{cav} . Therefore we write $X_{\text{mean}} \in B$ and $X_{\text{cav}} \in C$ in place of $X \in B$ and $X \in C$, respectively.

With this notation in place, we have

$$\begin{aligned} \mu(A) &= \frac{1}{Z_\beta} \sum_{X_{\text{cav}} \in C} \sum_{X_{\text{mean}} \in B} \exp(\mathcal{H}_\beta(X)) \\ &\leq \frac{1}{Z_\beta} \sum_{X_{\text{cav}} \in C} \sum_{X_{\text{mean}} \in B} \exp(\mathcal{H}_\beta^{\text{mean}}(X; S)) \exp(O(nr\eta + r^2) + \Gamma(p^*) + \bar{\Delta}(\mathbf{q}_S, p^*)) \\ &= \frac{H_{\text{cav}}(\mathbf{q}_S)}{Z_\beta} \sum_{X_{\text{mean}} \in B} \exp(\mathcal{H}_\beta^{\text{mean}}(X; S)) \exp(O(nr\eta + r^2) + \Gamma(p^*) + \bar{\Delta}(\mathbf{q}_S, p^*)) \\ (31) \quad &\leq \frac{Z_\beta^{\text{mean}}}{Z_\beta} \exp\left(O(nr\eta + r^2) + 2n \sum_{u \in S} \left[L_\beta(\bar{q}_u) - L_\beta(p^*) - \zeta_\beta(|\bar{q}_u - p^*|)\right]\right), \end{aligned}$$

where

$$\Gamma(p^*) := \sum_{i=0}^K \beta_i nr |V_i| (p^*)^{|E_i|},$$

$$\bar{\Delta}(\mathbf{q}_S, p^*) := 2n \sum_{u \in S} \left[L_\beta(\bar{q}_u) + I(\bar{q}_u) - L_\beta(p^*) - I(p^*) - \zeta_\beta(|\bar{q}_u - p^*|)\right],$$

and

$$Z_\beta^{\text{mean}} := \sum_{X_{\text{mean}} \in B} \exp(\mathcal{H}_\beta^{\text{mean}}(X; S)) \exp(-2n|S|I(p^*) + \Gamma(p^*)).$$

In the second step of (31) we have applied the cavity decomposition from Lemma 8.7. In the third step, we have used the fact that $|C| = H_{\text{cav}}(\mathbf{q}_S)$ and in the fourth step, we have used Lemma 8.8 to upper bound this count.

Now, note that under the condition $p^* < 1 - |S|/n$, there exists an admissible restricted degree $\bar{q} \in \{0, 1/n, \dots, 1 - |S|/n\}$ such that $|\bar{q} - p^*| \leq 1/n$. Denote by $\bar{\mathbf{q}}$ the degree tuple with $\bar{\mathbf{q}}_u = \bar{q}$ for every $u \in S$. Clearly, $\mu(B) \geq \mu(A(S, \bar{\mathbf{q}}, p^*, \eta))$. Repeating the calculation in Equation (31), but with corresponding lower bounds instead of upper bounds, we conclude that

$$(32) \quad \mu(B) \geq \frac{Z_\beta^{\text{mean}}}{Z_\beta} \exp(-O(nr\eta + r^2 \log(\frac{n}{r}) + r \log n) + \Delta(\bar{\mathbf{q}}, p^*)),$$

where

$$\Delta(\mathbf{q}_S, p^*) = 2n \sum_{u \in S} \left[L_\beta(\bar{q}_u) - L_\beta(p^*) - C_\beta |\bar{q}_u - p^*| \right].$$

Using the properties of the Shannon entropy (i.e, $\sup_{p \in [0, 1-1/n]} |H(p) - H(p+1/n)| = |H(0) - H(1/n)| \leq n^{-1}(1 + \log n)$), we have that

$$\sup_{p \in [0, 1-1/n]} \left| L_\beta(p) - L_\beta\left(p + \frac{1}{n}\right) \right| \leq C_\beta \frac{\log n}{n}$$

for some positive constant C_β . This implies that $\Delta(\bar{\mathbf{q}}, p^*) \leq C_\beta r \log(n)$. Plugging this into the lower bound on $\mu(B)$ in Equation (32), and then combining it with the upper bound on $\mu(A)$ in Equation (31), we obtain the claim. \square

From this result we derive the following corollary, which establishes that $\sup_{u \in [n]} |p_u(X) - p^*|$ must be close to zero with high probability under the measure $\mu(\cdot | B_\eta^\square(p^*))$. Recall that this corollary was stated and used in Section 6.1 in order to prove Theorem 4.5, which is the important component behind the proof of Theorem 3.1. Observe that in the statement of Theorem 8.9, the term in the exponent, $L_\beta(\bar{q}_u) - L_\beta(p^*) - \zeta_\beta |\bar{q}_u - p^*| < -\delta_0 < 0$ whenever $|\bar{q}_u - p^*|$ is large. Therefore, this the event where $|\bar{q}_u - p^*|$ is large incurs an exponentially small probability. We refer to Section D for its complete proof.

COROLLARY 6.3. Suppose $p^* \in U_\beta$ and let $\epsilon > 0$ be an arbitrary fixed constant. Then, for any $0 < \eta < c(\beta, \epsilon)$ and $n > n_0(\beta, \epsilon, \eta)$, we have:

$$\mu\left(\sup_{u \in [n]} |p_u(X) - p^*| \leq \epsilon \mid B_\eta^\square(p^*)\right) \geq 1 - \exp(-\Omega_{\beta, \epsilon}(n)).$$

Recall that in Lemma 7.1, we showed that whenever X is close to the constant graphon p^* , then *most* of the degrees concentrate close to p^* . In the result below we show that when $p_u(X)$ is close to p^* uniformly for *every* vertex u (as shown in Corollary 6.3) and X is close to the constant graphon p^* , most of the degrees $p_{uv}(X)$ concentrate close to $(p^*)^2$. Recall that this lemma was stated in Section 6, where it was used to prove Theorem 4.5.

COROLLARY 6.4. Suppose that $p^* \in U_\beta$. Given arbitrary $\epsilon, \delta > 0$, suppose $\eta < c(\beta, \epsilon, \delta)$ and $n > n_0(\delta, \epsilon, \beta, \eta)$. Then, for every $u \in [n]$, there exists a random set $S_u \subseteq [n] \setminus \{u\}$ such that $|S_u| \leq \delta n$ and

$$\mu \left(\sup_{u \in [n]} \sup_{v \in S_u^c} |p_{uv}(X) - (p^*)^2| \leq \epsilon \left| B_\eta^\square(p^*) \right| \right) \geq 1 - \exp(-\Omega_{\beta, \epsilon}(n)).$$

The corollary states that with high probability, for every u , most $p_{uv}(X)$ are such that $p_{uv}(X) \approx (p^*)^2$. The proof of this fact is essentially similar to the proof of Lemma 7.1. That is, we show that for any u , if $|p_{uv}(X) - p^* p_u(X)| > \epsilon$ for δn number of vertices v , then X violates the assumption that $\delta_\square(X, p^*) < \eta$. We then apply Corollary 6.3 to conclude that $p^* p_u(X)$ is uniformly close to $(p^*)^2$ with high probability for every u . We refer to Appendix D.1 for the full proof.

9. Proof of Theorem 3.3. Before proceeding with the proof of Theorem 3.3, we will establish generalizations of [1, Lemma 12 and Lemma 17]. Therefore, we will not instantiate to the model parameters given in Example 3.2 but consider a general ERGM with parameter β . We need to treat the vertex 1 separately from the other vertices. Following the notation preceding Lemma 12 in [1], we define for some fixed, finite set of graphs \mathcal{G} :

$$\bar{r}_{\max}(X) := \max \left(\max_{u: u \neq 1} p_u(X), \sup_{\substack{e=(u,v): u,v \neq 1 \\ G \in \mathcal{G}}} r_G(X, e) \right)$$

and

$$\bar{r}_{\min}(X) := \min \left(\min_{u: u \neq 1} p_u(X), \inf_{\substack{e=(u,v): u,v \neq 1 \\ G \in \mathcal{G}}} r_G(X, e) \right).$$

Here, we consider the evolution of the vertices $2, \dots, n-1$ when they are close to $G(n-1, p^*)$ in terms of the subgraph counts and the edges connecting vertex 1 are arbitrary. Notice that we have included the degrees $p_u(X)$ here in addition to $r_G(X, e)$, which will be useful to us later in the proof. The lemma below follows from a rewriting of the proof of Lemma 17 in [1], by noting that the edges connected to vertex 1 do not influence the evolution of $\Delta_G^e(X)$ in the leading order term as considered in [1, Lemma 12 and Lemma 14] and a straightforward tweak to also consider $p_u(X)$. Therefore, we skip the proof.

LEMMA 9.1. Suppose $\mathcal{G} = \mathbb{G}_L$ (the set of all graphs with at-most L vertices) and let $\epsilon > 0$ be a small enough constant independent of n . Suppose $p^* \in U_\beta$ and let X_0, X_1, \dots are drawn from the Glauber dynamics with respect to the measure μ . For some large enough $L \in \mathbb{N}$, independent of n , if X_0 is such that $p^* - \epsilon \leq \bar{r}_{\min}(X_0) \leq \bar{r}_{\max}(X_0) \leq p^* + \epsilon$, then for some α depending only on β, ϵ, L , we have

$$\mathbb{P} \left(\sup_{t \leq e^{\alpha n}} \bar{r}_{\max}(X_t) \geq p^* + 2\epsilon \right) \leq \exp(-\Omega_{\beta, L, \epsilon}(n))$$

and

$$\mathbb{P} \left(\inf_{t \leq e^{\alpha n}} \bar{r}_{\min}(X_t) \leq p^* - 2\epsilon \right) \leq \exp(-\Omega_{\beta, L, \epsilon}(n)).$$

We now instantiate our discussion to the case of the exponential random graph model defined in Example 3.2 and use the notation established in this example. Recall $p_1(X)$ and $p_{1u}(X)$. Define

$$p_{\max}^{(1)}(X) = \max \left(p_1(X), \max_{\substack{u \neq 1 \\ u \in [n]}} \frac{p_{1u}(X)}{p_1^*} \right) \quad \text{and} \quad p_{\min}^{(1)}(X) = \min \left(p_1(X), \min_{\substack{u \neq 1 \\ u \in [n]}} \frac{p_{1u}(X)}{p_1^*} \right).$$

LEMMA 9.2. Consider the same setting as Lemma 9.1 instantiated to the parameter β given in Example 3.2, with $p^* = p_1^*$. Given $\epsilon_1 > 0$, we can pick ϵ in Lemma 9.1 small enough such that the following holds: Suppose $q^* - \epsilon_1 \leq p_{\min}^{(1)}(X_0) \leq p_{\max}^{(1)}(X_0) \leq q^* + \epsilon_1$, then for some α depending only on $\beta, \epsilon, \epsilon_1, L$, we have

$$\mathbb{P}\left(\sup_{t \leq e^{\alpha n}} p_{\max}^{(1)}(X_t) \geq q^* + 2\epsilon_1\right) \leq \exp(-\Omega_{\beta, L, \epsilon}(n))$$

and

$$\mathbb{P}\left(\inf_{t \leq e^{\alpha n}} p_{\min}^{(1)}(X_t) \leq q^* - 2\epsilon_1\right) \leq \exp(-\Omega_{\beta, L, \epsilon}(n)).$$

PROOF. Let $N := \binom{n}{2}$. Recall the function g defined in Example 3.2 and ϕ_β as defined in Section 2. It is easy to show using similar techniques as in [1, Lemma 12] that

$$-\frac{p_1(X_t)}{N} + \frac{g(p_{\min}^{(1)})}{N} \leq \mathbb{E}[p_1(X_{t+1}) - p_1(X_t) | X_t] \leq -\frac{p_1(X_t)}{N} + \frac{g(p_{\max}^{(1)})}{N}.$$

Similarly, for every $u \in [n]$ and $u \neq 1$, denoting $\bar{r}_{\max}(X_t), \bar{r}_{\min}(X_t)$ by $\bar{r}_{\max}, \bar{r}_{\min}$ respectively,

$$\mathbb{E}[p_{1u}(X_{t+1}) - p_{1u}(X_t) | X_t] \leq -\frac{2p_{1u}(X_t)}{N} + \frac{g(p_{\max}^{(1)})\bar{r}_{\max}}{N} + \frac{p_{\max}^{(1)}\phi(\bar{r}_{\max})}{N}$$

and

$$\mathbb{E}[p_{1u}(X_{t+1}) - p_{1u}(X_t) | X_t] \geq -\frac{2p_{1u}(X_t)}{N} + \frac{g(p_{\min}^{(1)})\bar{r}_{\min}}{N} + \frac{p_{\min}^{(1)}\phi(\bar{r}_{\min})}{N}.$$

Now, notice that by Lemma 9.1, $\bar{r}_{\max}(X_t)/p_1^* \leq 1 + 2\frac{\epsilon}{p_1^*}$ and $\bar{r}_{\min}(X_t)/p_1^* \geq 1 - 2\frac{\epsilon}{p_1^*}$ with probability at-least $\exp(-\Omega(n))$ whenever $t \leq \exp(\alpha n)$. Therefore, we can consider the evolution of $\frac{p_{1u}(X_t)}{p_1^*}$ akin to the evolution of $p_1(X_t)$ with $g(x)$ replaced by $\frac{g(x)+x}{2}$. Notice that the functions g and $\frac{1}{2}(g(x)+x)$ play the role of $\phi()$ in the proof of [1, Lemma 12] and satisfy the relationship $g(q^*) = q^*$ and $g'(q^*) < 1$ (same for $\frac{g(x)+x}{2}$). This allows us to conclude the statement of the lemma with minor modifications to the proof of Lemma 17 in [1]. \square

PROOF OF THEOREM 3.3. Let the initial state X_0 be sampled as in the theorem statement.

1. Recall $\delta_{\square}^{S, p^*}$ given in Definition 8.5. Consider this with $S = \{1\}$. By Lemma 8.6, in order to show that $X_t \in B_{\eta}^{\square}(p^*)$ it is sufficient to show that $\delta_{\square}^{1, p^*}(X_t, p^*) \leq \eta/4$ with high probability. With similar arguments as in the proof of Lemma 4.11 with r_{\min}, r_{\max} replaced with $\bar{r}_{\min}, \bar{r}_{\max}$ we conclude that with probability at-least $1 - T \exp(-\Omega_{\eta, \beta}(n))$, we have that every point in the trajectory $X_0, X_1, \dots, X_T \in B_{\eta}^{\square}(p^*)$. Using Lemma 9.2 and the result above we conclude the statement.
2. This follows from a straightforward application of Theorem 8.9 along with Theorem 1.3 and the fact that p_1^* is the unique global maximizer of U_β .
3. This follows from the same considerations as the proof of Item 2.

\square

Acknowledgements and Funding. G.B. and D.N. gratefully acknowledge the hospitality of the Simons Institute for Theoretical Computer Science during Fall 2020 and Fall 2021. This work was supported in part by NSF CAREER award CCF-1940205 and NSF Award DMS-2022448.

REFERENCES

- [1] BHAMIDI, S., BRESLER, G. and SLY, A. (2011). Mixing time of exponential random graphs. *Annals of Applied Probability* **21** 2146–2170.
- [2] BORGES, C., CHAYES, J., LOVÁSZ, L., SÓS, V. T., SZEGEDY, B. and VESZTERGOMBI, K. (2006). Graph limits and parameter testing. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing* 261–270.
- [3] BRESLER, G. and NAGARAJ, D. (2018). Optimal single sample tests for structured versus unstructured network data. In *Conference On Learning Theory* 1657–1690. PMLR.
- [4] BURDA, Z., JURKIEWICZ, J. and KRZYWICKI, A. (2004). Network transitivity and matrix models. *Physical Review E* **69** 026106.
- [5] CHATTERJEE, S. (2007). Stein’s method for concentration inequalities. *Probability theory and related fields* **138** 305–321.
- [6] CHATTERJEE, S. and DIACONIS, P. (2013). Estimating and understanding exponential random graph models. *The Annals of Statistics* **41** 2428–2461.
- [7] ELDAN, R. and GROSS, R. (2018). Exponential random graphs behave like mixtures of stochastic block models. *The Annals of Applied Probability* **28** 3698–3735.
- [8] FIENBERG, S. E. (2010). Introduction to papers on the modeling and analysis of network data. *The Annals of Applied Statistics* 1–4.
- [9] FIENBERG, S. E. (2010). Introduction to papers on the modeling and analysis of network data—II. *The Annals of Applied Statistics* **4** 533–534.
- [10] FRANK, O. and STRAUSS, D. (1986). Markov graphs. *Journal of the american Statistical association* **81** 832–842.
- [11] GANGULY, S. and NAM, K. (2019). Sub-critical Exponential random graphs: concentration of measure and some applications. *To appear in Transactions of the AMS. arXiv preprint arXiv:1909.11080*.
- [12] GELMAN, A. and MENG, X.-L. (1998). Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical science* 163–185.
- [13] GEYER, C. J. and THOMPSON, E. A. (1992). Constrained Monte Carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society: Series B (Methodological)* **54** 657–683.
- [14] GHEISSARI, R. and SINCLAIR, A. (2021). Low-temperature Ising dynamics with random initializations. *arXiv preprint arXiv:2106.11296*.
- [15] GUO, H. and JERRUM, M. (2017). Random cluster dynamics for the Ising model is rapidly mixing. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms* 1818–1827. SIAM.
- [16] HOLLAND, P. W. and LEINHARDT, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the american Statistical association* **76** 33–50.
- [17] JERRUM, M. and SINCLAIR, A. (1993). Polynomial-time approximation algorithms for the Ising model. *SIAM Journal on computing* **22** 1087–1116.
- [18] KOU, S., ZHOU, Q. and WONG, W. H. (2006). Equi-energy sampler with applications in statistical inference and statistical mechanics. *The annals of Statistics* **34** 1581–1619.
- [19] LEVIN, D. A., LUCZAK, M. J. and PERES, Y. (2010). Glauber dynamics for the mean-field Ising model: cut-off, critical power law, and metastability. *Probability Theory and Related Fields* **146** 223–265.
- [20] LEVIN, D. A. and PERES, Y. (2017). *Markov chains and mixing times* **107**. American Mathematical Soc.
- [21] LINDSEY, M., WEARE, J. and ZHANG, A. (2021). Ensemble Markov chain Monte Carlo with teleporting walkers. *arXiv preprint arXiv:2106.02686*.
- [22] LUBETZKY, E. and SLY, A. (2021). Fast initial conditions for Glauber dynamics. *Probability Theory and Related Fields* **181** 647–667.
- [23] MUKHERJEE, S. and XU, Y. (2013). Statistics of the two-star ERGM. *arXiv preprint arXiv:1310.4526*.
- [24] PARK, J. and NEWMAN, M. E. (2004). Solution of the two-star model of a network. *Physical Review E* **70** 066146.
- [25] PARK, J. and NEWMAN, M. E. (2005). Solution for the properties of a clustered network. *Physical Review E* **72** 026136.
- [26] RADIN, C. and YIN, M. (2013). Phase transitions in exponential random graphs. *The Annals of Applied Probability* 2458–2471.
- [27] REINERT, G. and ROSS, N. (2019). Approximating stationary distributions of fast mixing Glauber dynamics, with applications to exponential random graphs. *The Annals of Applied Probability* **29** 3201–3229.
- [28] SINCLAIR, A. and JERRUM, M. (1989). Approximate counting, uniform generation and rapidly mixing Markov chains. *Information and Computation* **82** 93–133.
- [29] SWENDSEN, R. H. and WANG, J.-S. (1987). Nonuniversal critical dynamics in Monte Carlo simulations. *Physical review letters* **58** 86.

- [30] ULLRICH, M. (2014). Swendsen–Wang Is Faster than Single-Bond Dynamics. *SIAM Journal on Discrete Mathematics* **28** 37–48.
- [31] WASSERMAN, S. and FAUST, K. (1994). Social network analysis: Methods and applications.
- [32] YIN, M. and ZHU, L. (2017). Asymptotics for sparse exponential random graph models. *Brazilian Journal of Probability and Statistics* **31** 394–412.

APPENDIX A: PROOFS OF TECHNICAL LEMMAS

A.1. Proof Sketch of Lemma 4.8.

$$\frac{\mu(\eta \leq \delta_{\square}(X, p^*) \leq \eta + \delta)}{\mu(\delta_{\square}(X, p^*) \leq \eta)} \leq C(\eta, \delta) \exp(-c(\eta, \delta)n^2).$$

Note that by Theorem 1.3, we know that

$$\mu(\eta \leq \delta_{\square}(X, p^*) \leq \eta + \delta) \leq C_1(\eta, \delta) \exp(-c_1(\eta, \delta)n^2)$$

This is true since X in this set is not within an η neighborhood of \tilde{M}_β whenever δ is small enough as assumed. The proof follows once we demonstrate that:

$$(33) \quad \liminf_{n \rightarrow \infty} \frac{\log \mu_n(\delta_{\square}(X, p^*) \leq \eta)}{n^2} \geq 0$$

Here, we denote μ by μ_n in order to explicate the dependence on n . Taking the partition function to be $Z_\beta = \exp(n^2\psi_n)$, we conclude from Theorem 4.1 in [6, Theorem 4.1] that

$$\lim_{n \rightarrow \infty} \psi_n = L_\beta(p^*).$$

Let $Z_\beta(\eta, p^*) := \sum_{X: \delta_{\square}(X, p^*) \leq \eta} \exp(\mathcal{H}_\beta(X)) = \exp(n^2\psi_n(\eta, p^*))$. Therefore, showing Equation (33) reduces to showing that:

$$\liminf_{n \rightarrow \infty} \psi_n(\eta, p^*) \geq \lim_{n \rightarrow \infty} \psi_n = L_\beta(p^*).$$

We now follow the proof of [6, Theorem 3.1] with some modifications. Note that the Hamiltonian $\mathcal{H}_\beta(X) := \sum_{i=0}^K n^2 \beta_i N_i(X)$ can be viewed as a function on the graphon counterpart of X with $\mathcal{H}_\beta(X)/n^2$ being a continuous function over the space of graphons. This is identified with the function T in [6].

Now, replacing \tilde{U}^a with $T^{-1}((a, a + \epsilon)) \cap \{\tilde{h} \in \tilde{\mathcal{W}} : \delta_{\square}(\tilde{h}, p^*) < \eta\}$, we conclude that:

$$\liminf_n \psi_n(\eta, p^*) \geq \sup_{\delta_{\square}(\tilde{h}, p^*) < \eta} T(\tilde{h}) - I(\tilde{h}).$$

Where the function I is extended to the space of graphons via a pointwise application like in [6]. By [6, Theorem 4.1] and its proof, it is clear that $T(\tilde{h}) - I(\tilde{h})$ is maximized at the constant p^* graphon and the corresponding value is $L_\beta(p^*)$. Therefore, from the discussion above, we conclude:

$$\liminf_n \psi_n(\eta, p^*) = L_\beta(p^*) = \lim_n \psi_n$$

This establishes the lemma. □

A.2. Proof of Lemma 7.1. Fix any $p \in [0, 1]$ and $\delta > 0$ and consider the sets $S^+ := \{u : p_u(X) - p > 2\delta_{\square}(\tilde{X}, p\mathbf{1})/\delta\}$ and $S^- := \{u : p_u(X) - p < -2\delta_{\square}(\tilde{X}, p\mathbf{1})/\delta\}$. By the definition of the cut-metric it follows that

$$\delta_{\square}(\tilde{X}, p\mathbf{1}) \geq \frac{1}{n} \sum_{u \in S^+} p_u(X) - p \geq \frac{2|S^+|\delta_{\square}(\tilde{X}, p\mathbf{1})}{n\delta}.$$

Thus, $|S^+| \leq \delta n/2$ and similarly $|S^-| \leq \delta n/2$. Therefore, $|S^+ \cup S^-| \leq \delta n$, which allows us to conclude the first inequality by taking $S = S^+ \cup S^-$. The second inequality follows by directly applying Theorem 1.3 to Equation (19). □

A.3. Proof of Lemma 7.2. We note that when considering distance from any constant graphon $p\mathbf{1}$ we have $\delta_{\square}(\tilde{X}, p\mathbf{1}) = \delta_{\square}(f^X, p\mathbf{1})$, since measure-preserving operators do not affect the constant graphon. Now, notice that since $f^Y(x_1, x_2) - p \leq f^X(x_1, x_2) - p \leq f^Z(x_1, x_2) - p$, for any Borel measurable sets $S, T \subseteq [0, 1]$ we have

$$\int_{S \times T} (f^X(x_1, x_2) - p) dx_1 dx_2 \leq \int_{S \times T} (f^Z(x_1, x_2) - p) dx_1 dx_2 \leq \delta_{\square}(\tilde{Z}, p).$$

Similarly,

$$\int_{S \times T} (f^X(x_1, x_2) - p) dx_1 dx_2 \geq \int_{S \times T} (f^Y(x_1, x_2) - p) dx_1 dx_2 \geq -\delta_{\square}(\tilde{Y}, p).$$

These together establish that

$$\left| \int_{S \times T} (f^X(x_1, x_2) - p) dx_1 dx_2 \right| \leq \max(\delta_{\square}(\tilde{Y}, p), \delta_{\square}(\tilde{Z}, p)).$$

Taking the supremum over S, T yields the lemma. \square

A.4. Proof of Lemma 7.3. In case $d_i < |E|$ for some i , we start with Young's product inequality, which states that for every $a, b \geq 0$ and $\alpha, \beta > 0$ such that $\alpha^{-1} + \beta^{-1} = 1$, we have $ab \leq \alpha^{-1}a^{\alpha} + \beta^{-1}b^{\beta}$. We take $a = q^{d_i}$, $b = p^{|E|-d_i}$, $\alpha = |E|/d_i$ and $\beta = |E|/(|E| - d_i)$. In case $d_i = |E|$, we check that the inequalities below hold trivially. Hence

$$\begin{aligned} \sum_{i=1}^{|V|} q^{d_i} p^{|E|-d_i} &\leq \sum_{i=1}^{|V|} \frac{d_i}{|E|} q^{|E|} + \frac{|E| - d_i}{|E|} p^{|E|} \\ &= \frac{\sum_{i=1}^{|V|} d_i}{|E|} q^{|E|} + \frac{|V||E| - \sum_{i=1}^{|V|} d_i}{|E|} p^{|E|} \\ (34) \quad &= 2q^{|E|} + (|V| - 2)p^{|E|}. \end{aligned}$$

In the last step, we have used the fact that for any finite simple graph G , $\sum_{i=1}^{|V|} d_i = 2|E|$. Equality when $G = G_0$ follows by a straightforward calculation.

Now suppose that $G \neq G_0$. Then, it is easy to show that there exists a vertex j such that $d_j < |E|$. We note that Young's product inequality is strict whenever $a^{\alpha} \neq b^{\beta}$. For the choice of a, b, α, β above, this condition means $p \neq q$. Now, consider the function: $f(p, q) = -q^{d_j} p^{|E|-d_j} + \frac{d_j}{|E|} q^{|E|} + \frac{|E|-d_j}{|E|} p^{|E|}$. This is continuous over the set $[0, 1]^2$. Define $A_{\delta} := \{(p, q) \in [0, 1]^2 : |p - q| \geq \delta\}$. Clearly, A_{δ} is a compact set for every $\delta \geq 0$. Define $\zeta(\delta) = \inf_{(p, q) \in A_{\delta}} f(p, q)$. It is clear from the strictness of the Young's inequality that $f(p, q) > 0$ for every $(p, q) \in A_{\delta}$ whenever $\delta > 0$. Therefore, we conclude by compactness of A_{δ} and continuity of f that $\zeta(\delta) > 0$ whenever $\delta > 0$. The continuity of ζ follows from the continuity of f . Therefore, we conclude that there exists ζ as in the statement of the lemma such that:

$$-C|p - q| \leq q^{d_j} p^{|E|-d_j} - \frac{d_j}{|E|} q^{|E|} - \frac{|E| - d_j}{|E|} p^{|E|} \leq -\zeta(|p - q|).$$

The inequality above holds with $\zeta = 0$ for every i , even when $d_i = |E|$. This allows us to sum the inequality above and conclude the result. \square

A.5. Proof of Lemma 8.2. Since the proof is elementary, we only provide a brief sketch. The statement $N_G^0(X, u) \geq N_G(X, u)$ follows from the fact that $A_u^k = \bigcup_{l=1}^k A_u^{k,l}$ and the union bound. Now, note that $0 \leq \sum_{l=1}^k \mathbb{1}(A_u^{k,l}) - \mathbb{1}(A_u^k) \leq k$ and the sum is non zero only when the event $A_u^{k,l} \cap A_u^{k,m}$ holds for some $l \neq m, l, m \in [k]$. Noting that under the uniform measure over $[0, 1]^k$ the measure of $A_u^{k,l} \cap A_u^{k,m}$ is n^{-2} and using the union bound, we conclude the result. \square

A.6. Proof of Lemma 8.3. Suppose d_i is the degree of vertex $i \in [k]$. In light of Lemma 8.2 we may replace $N_G(X, u)$ in the lemma statement by $N_G^0(X; u)$ and then by considering a specific term in the sums we see that it is sufficient to prove that

$$\left| \int_{[0,1]^k} \mathbb{1}(A_u^{k,l}) \prod_{(i,j) \in E(G)} f^X(x_i, x_j) \prod_{i=1}^k dx_i - \frac{p_u(X)^{d_l} p^{|E(G)|-d_l}}{n} \right| \leq |E(G)| \frac{\delta_{\square}(\tilde{X}, p)}{n}.$$

Notice that

$$(35) \quad \begin{aligned} & \int_{[0,1]^k} \mathbb{1}(A_u^{k,l}) \prod_{(i,j) \in E(G)} f^X(x_i, x_j) \prod_{i=1}^k dx_i \\ &= \int_{[0,1]^k} \mathbb{1}(A_u^{k,l}) \prod_{(l,i) \in E(G)} f^X(x_l, x_i) \prod_{\substack{(i,j) \in E(G) \\ i,j \neq l}} f^X(x_i, x_j) \prod_{i=1}^k dx_i. \end{aligned}$$

A simple computation shows that

$$(36) \quad \frac{p_u(X)^{d_l} p^{|E(G)|-d_l}}{n} = \int_{[0,1]^k} \mathbb{1}(A_u^{k,l}) \left(\prod_{i:(l,i) \in E(G)} f^X(x_l, x_i) \right) p^{|E(G)|-d_l} \prod_{i=1}^k dx_i.$$

Therefore,

$$(37) \quad \begin{aligned} & \left| \int_{[0,1]^k} \mathbb{1}(A_u^{k,l}) \prod_{(i,j) \in E(G)} f^X(x_i, x_j) \prod_{i=1}^k dx_i - \frac{p_u(X)^{d_l} p^{|E(G)|-d_l}}{n} \right| \\ &= \left| \int_{[0,1]^k} \mathbb{1}(A_u^{k,l}) \prod_{(l,i) \in E(G)} f^X(x_l, x_i) \left[\prod_{\substack{(i,j) \in E(G) \\ i,j \neq l}} f^X(x_i, x_j) - p^{|E(G)|-d_l} \right] \prod_{i=1}^k dx_i \right|. \end{aligned}$$

Following the proof of [2, Lemma 4.4] with minor modifications, consider any ordering among the set of edges $(i,j) \in E(G)$ such that $i, j \neq l$ and index these ordered edges by $(i_1, j_1), \dots, (i_h, j_h)$ where $h = |E(G)| - d_l$. Then

$$\begin{aligned} & \left[\prod_{\substack{(i,j) \in E(G) \\ i,j \neq l}} f^X(x_i, x_j) - p^{|E(G)|-d_l} \right] \\ &= \sum_{r=0}^{|E(G)|-d_l-1} p^r \prod_{m=r+1}^{|E(G)|-d_l} f^X(x_{i_m}, x_{j_m}) - p^{r+1} \prod_{m=r+2}^{|E(G)|-d_l} f^X(x_{i_m}, x_{j_m}). \end{aligned}$$

Now, we use the above decomposition in Equation (37) and consider the terms in the summation one by one. We then follow the technique used in the proof of [2, Lemma 4.4] along with the fact that $A_u^{k,l}$ depends only on x_u and the fact that the measure of the event $A_u^{k,l}$ under the uniform measure over $[0,1]^k$ is $1/n$ to conclude the result. \square

A.7. Proof of Lemma 8.4. Only in this proof, we will take the probability space to be $[0,1]^k$ equipped with the Borel sigma algebra and the uniform measure P .

First, note that by the union bound, $\sum_{u \in S} N_G(X; u) \geq N_G(X; S)$. Now, almost surely

$$\sum_{u \in S} \mathbb{1}(A_u^k) \leq k,$$

since there are at most k vertices in the graph G . We conclude that almost surely

$$\sum_{u \in S} \mathbb{1}(A_u^k) - \mathbb{1}(A_S^k) \leq k - 1.$$

Now let B_S^k be the event that $\sum_l \mathbb{1}(A_l^k) \neq \mathbb{1}(A_S^k)$. This can happen only when two events $A_u^{k,i}$ and $A_v^{k,j}$ hold simultaneously for some $i, j \in [k]$, $u, v \in S$, $i \neq j$ and $u \neq v$. Therefore, we have

$$B_S^k = \bigcup_{\substack{u, v \in S \\ u \neq v}} \bigcup_{\substack{i, j \in [k] \\ i \neq j}} A_u^{k,i} \cap A_v^{k,j}.$$

By the union bound,

$$P(B_S^k) \leq |S|^2 k^2 P(A_l^{k,i} \cap A_m^{k,j}) = \frac{|S|^2 k^2}{n^2}.$$

Now combining the considerations above, we have

$$\begin{aligned} & \sum_{l \in S} N_G(X; l) - N_G(X; S) \\ &= \int_{[0,1]^k} \left[-\mathbb{1}(A_S^k) + \sum_{l \in S} \mathbb{1}(A_l^k) \right] \prod_{(i,j) \in E(G)} f^X(x_i, x_j) \prod_{i=1}^k dx_i \\ &\leq (k-1) \int_{[0,1]^k} \mathbb{1}(B_S^k) \prod_{(i,j) \in E(G)} f^X(x_i, x_j) \prod_{i=1}^k dx_i \\ (38) \quad &\leq k P(B_S^k) \leq \frac{|S|^2 k^3}{n^2}. \end{aligned}$$

The lemma statement follows. \square

A.8. Proof of Lemma 8.6. In this proof alone, we will abuse notation to denote the set $\cup_{u \in S} [\frac{u-1}{n}, \frac{u}{n}] \subseteq [0,1]$ also by S (and similarly for S^C). Since we are considering the cut-metric between \tilde{X} and a constant graphon, we can write

$$\begin{aligned} \delta_{\square}^{S,p^*}(\tilde{X}, p^*) &= \sup_{A, B \subset [0,1]} \left| \int_{A \times B} [f^{X,S,p^*}(x_1, x_2) - p^*] dx_1 dx_2 \right| \\ &= \sup_{A, B \subset [0,1]} \left| \int_{A \cap S^C \times B \cap S^C} [f^X(x_1, x_2) - p^*] dx_1 dx_2 \right| \\ (39) \quad &= \sup_{A, B \subset S^C} \left| \int_{A \times B} [f^X(x_1, x_2) - p^*] dx_1 dx_2 \right|. \end{aligned}$$

From the last equality above, we conclude that $\delta_{\square}^{S,p^*}(\tilde{X}, p^*) \leq \delta_{\square}(\tilde{X}, p^*)$. To conclude the lower bound, note that

$$\delta_{\square}(\tilde{X}, p) \leq \delta_{\square}(\tilde{X}, \tilde{X}^{S,p^*}) + \delta_{\square}(\tilde{X}^{S,p^*}, p) = \delta_{\square}(\tilde{X}, \tilde{X}^{S,p^*}) + \delta_{\square}^{S,p^*}(\tilde{X}, p^*).$$

It is now easy to show that $\delta_{\square}(\tilde{X}, \tilde{X}^{S,p^*}) \leq 1 - |S^C|^2/n^2$, which when combined with the last display above proves the lower bound. \square

A.9. Proof of Lemma 8.8. By Stirling's formula, we have that for any $k \in \mathbb{N}$ and $p \in [0, 1]$ such that kp is an integer,

$$\frac{\exp(-2kI(p))}{\sqrt{2k}} \leq \binom{k}{kp} \leq \exp(-2kI(p)).$$

A counting argument shows that $N_{\text{cav}}(\mathbf{q}_S) = \prod_{u \in S} \binom{n-|S|}{q_u n}$. For the upper bound, note that

$$(40) \quad \prod_{u \in S} \binom{n-|S|}{q_u n} \leq \prod_{u \in S} \binom{n}{q_u n} \leq \exp\left(-2n \sum_{u \in S} I(q_u)\right).$$

Now, for the lower bounds, we note that whenever $q_u \leq q \leq 1$ and $q \geq 1/2$: $|2qI(q_u/q) - 2I(q_u)| \leq (2 + \log(1/(1-q)))(1-q)$. Taking $q = (n-|S|)/n = 1 - r/n$ below, we have

$$(41) \quad \begin{aligned} \prod_{u \in S} \binom{n-|S|}{q_u n} &\geq \frac{\exp\left(-\sum_{u \in S} 2(n-|S|)I\left(\frac{nq_u}{n-|S|}\right)\right)}{(\sqrt{2n})^r} \\ &\geq \exp\left(-\sum_{u \in S} 2nI(q_u) - r^2 \left[4 + 2\log\left(\frac{n}{r}\right)\right] - \frac{r}{2} \log(2n)\right). \end{aligned}$$

□

APPENDIX B: PROOFS OF LEMMAS FROM SECTION 6

B.1. Proof of Lemma 6.1. Before proving the lemma, we derive an estimate for $\Delta_G^e(X)$.

Recall the event A_u^k in (21) in Section 8. Suppose the fixed graph G has the vertex set $[k]$. We now define for $u, v \in [n]$,

$$A_{uv}^{kG} := \{x \in [0, 1]^k : (\lfloor nx_i \rfloor, \lfloor nx_j \rfloor) \in \{(u, v), (v, u)\} \text{ for some } (i, j) \in E(G)\}.$$

For $i, j \in [k]$ define

$$A_{uv}^{kij} := \{x \in [0, 1]^k : (\lfloor nx_i \rfloor, \lfloor nx_j \rfloor) \in \{(u, v), (v, u)\}\}.$$

Now, the definition of homomorphism density yields

$$(42) \quad \begin{aligned} \Delta_G^e(X) &= N_G(X^{+e}) - N_G(X^{-e}) \\ &= \int \prod_{(i,j) \in E(G)} f^{X^{+e}}(x_i, x_j) \prod_{t=1}^k dx_t - \prod_{(i,j) \in E(G)} f^{X^{-e}}(x_i, x_j) \prod_{t=1}^k dx_t \\ &= \int \mathbb{1}(A_{uv}^{kG}) \prod_{(i,j) \in E(G)} f^{X^{+e}}(x_i, x_j) \prod_{t=1}^k dx_t. \end{aligned}$$

A computation similar to the proof of Lemma 8.2 shows that

$$\int \left| \mathbb{1}(A_{uv}^{kG}) - \sum_{(i,j) \in E(G)} \mathbb{1}(A_{uv}^{kij}) \right| \prod_{t=1}^k dx_t \leq \frac{E(G)^3}{n^3}.$$

Using this in Equation (42), we conclude that

$$\Delta_G^e(X) = \int \sum_{(i,j) \in E(G)} \mathbb{1}(A_{uv}^{kij}) \prod_{(a,b) \in E(G)} f^{X^{+e}}(x_a, x_b) \prod_{t=1}^k dx_t \pm O\left(\frac{E(G)^3}{n^3}\right)$$

$$(43) \quad = \sum_{(i,j) \in E(G)} \int \mathbb{1}(A_{uv}^{kij}) \prod_{\substack{(a,b) \in E(G) \\ (a,b) \neq (i,j)}} f^{X^{+e}}(x_a, x_b) \prod_{t=1}^k dx_t \pm O\left(\frac{E(G)^3}{n^3}\right).$$

We are now ready for the proof.

PROOF. We will prove the result by replacing $p_{uv}(X), p_u(X)$, and $\delta_{\square}(X, p^*)$ with $p_{uv}(X^{+e}), p_u(X^{+e})$, and $\delta_{\square}(X^{+e}, p^*)$, noting that $\delta_{\square}(X, X^{+e}) \leq 2/n^2$, $|p_{uv}(X) - p_{uv}(X^{+e})| \leq 1/n$, and $|p_u(X) - p_u(X^{+e})| \leq 1/n$.

We will use E and $E(G)$ interchangeably in this proof. For $(i, j) \in E(G)$, first consider the quantity

$$\begin{aligned} & \int \mathbb{1}(A_{uv}^{kij}) \prod_{\substack{(a,b) \in E \\ (a,b) \neq (i,j)}} f^{X^{+e}}(x_a, x_b) \prod_{t=1}^k dx_t \\ &= \int \mathbb{1}(A_{uv}^{kij}) \prod_{\substack{(i,l) \in E \\ l \neq j}} f^{X^{+e}}(x_i, x_l) \prod_{\substack{(j,l) \in E \\ l \neq i}} f^{X^{+e}}(x_j, x_l) \prod_{\substack{(a,b) \in E \\ a, b \notin \{i, j\}}} f^{X^{+e}}(x_a, x_b) \prod_{t=1}^k dx_t \\ &= (p^*)^{|E| - d_i - d_j + 1} \int \mathbb{1}(A_{uv}^{kij}) \prod_{\substack{(i,l) \in E \\ l \neq j}} f^{X^{+e}}(x_i, x_l) \prod_{\substack{(j,l) \in E \\ l \neq i}} f^{X^{+e}}(x_j, x_l) \prod_{t=1}^k dx_t \\ (44) \quad & \pm \frac{2|E(G)|\delta_{\square}(X^{+e}, p^*)}{n^2}. \end{aligned}$$

In the last step we have used a similar peeling argument as in Lemma 8.3. Recalling the sets A_u^{ki} from Section 8, we note that whenever $i \neq j$, $\mathbb{1}(A_{uv}^{kij}) = \mathbb{1}(A_u^{ki} \cap A_v^{kj}) + \mathbb{1}(A_u^{kj} \cap A_v^{ki})$. Now,

$$\begin{aligned} & \int \mathbb{1}(A_u^{ki} \cap A_v^{kj}) \prod_{\substack{(i,l) \in E \\ l \neq j}} f^{X^{+e}}(x_i, x_l) \prod_{\substack{(j,l) \in E \\ l \neq i}} f^{X^{+e}}(x_j, x_l) \prod_{t=1}^k dx_t \\ &= \int \mathbb{1}(A_u^{ki} \cap A_v^{kj}) \prod_{\substack{(i,l) \in E \\ l \neq j \\ l \notin E_{ij}(G)}} f^{X^{+e}}(x_i, x_l) \prod_{\substack{(j,l) \in E \\ l \neq i \\ l \in E_{ij}(G)}} f^{X^{+e}}(x_j, x_l) \times \\ & \quad \prod_{l \in E_{ij}(G)} f^{X^{+e}}(x_i, x_l) f^{X^{+e}}(x_j, x_l) \prod_{t=1}^k dx_t \\ (45) \quad &= \frac{1}{n^2} (p_u(X^{+e}))^{d_i - 1 - d_{ij}} (p_v(X^{+e}))^{d_j - 1 - d_{ij}} (p_{uv}(X^{+e}))^{d_{ij}}. \end{aligned}$$

In the last step, we have used the definitions of degrees p_u and p_{uv} given in Equations (4) and (5) in terms of integrals over f^X . Using this in Equation (44), and shortening $p_u(X_e^+), p_v(X_e^+)$ and $p_{uv}(X_e^+)$ to p_u, p_v and p_{uv} , we obtain

$$n^2 \int \mathbb{1}(A_{uv}^{kij}) \prod_{\substack{(a,b) \in E \\ (a,b) \neq (i,j)}} f^{X^{+e}}(x_a, x_b) \prod_{t=1}^k dx_t$$

$$\begin{aligned}
&= (p^*)^{|E|-d_i-d_j+1} (p_u)^{d_i-1-d_{ij}} (p_v)^{d_j-1-d_{ij}} (p_{uv})^{d_{ij}} \\
(46) \quad &+ (p^*)^{|E|-d_i-d_j+1} (p_v)^{d_i-1-d_{ij}} (p_u)^{d_j-1-d_{ij}} (p_{uv})^{d_{ij}} \pm 2|E(G)|\delta_{\square}(X^{\tilde{+}e}, p^*).
\end{aligned}$$

Using the condition $\sup_u |p_u(X) - p^*| < \epsilon$, Equation (43), and the equation above, we conclude the statement of the lemma. \square

B.2. Proof of Lemma 6.6. We first state a simple modification of [5, Theorem 1.5] below, which follows by essentially rewriting its proof with minor changes.

LEMMA B.1 (Modification of Theorem 1.5 in [5]). *Under the same notation as [5, Theorem 1.5], we suppose the same conditions as the original theorem hold, except condition (ii), where we replace the assumption $\Delta(X) \leq Bf(X) + C$, with*

1. $\Delta(X) \leq \alpha \mathbb{1}(A) + \gamma \mathbb{1}(A^C)$ for some event $A \in \sigma(X)$, $\alpha, \gamma \in \mathbb{R}^+$
2. $|f(X)| \leq M$ almost surely.

Then, we have

$$\mathbb{P}(|f(X)| > t) \leq \left(1 + \frac{\gamma \exp(\theta_0 M) \mathbb{P}(A^C)}{\alpha}\right) \inf_{\theta \in [0, \theta_0]} \exp\left(\frac{\theta^2 \alpha}{2} - \theta t\right).$$

PROOF. In the proof of [5, Theorem 1.5], in the display below Equation (7), we have for $\theta \in [0, \theta_0]$

$$\begin{aligned}
|m'(\theta)| &\leq |\theta| \mathbb{E} \exp^{\theta f(X)} \Delta(X) \\
&\leq |\theta| \alpha m(\theta) + |\theta| \gamma \exp(|\theta| M) \mathbb{P}(A^C) \\
&\leq |\theta| \alpha m(\theta) + |\theta| \gamma \exp(|\theta_0| M) \mathbb{P}(A^C).
\end{aligned}$$

In the second step we have used the hypothesis that $\Delta(X) \leq \alpha \mathbb{1}(A) + \gamma \mathbb{1}(A^C)$ and $|f(X)| \leq M$. Therefore

$$\frac{d}{d\theta} \log \left(m(\theta) + \frac{\gamma \exp(\theta_0 M) \mathbb{P}(A^C)}{\alpha} \right) \leq \theta \alpha.$$

The result then follows by an application of Gronwall's lemma and the Chernoff bound. \square

We will consider Glauber dynamics with respect to the measure $\mu(\cdot | B_{\eta}^{\square}(p^*))$ (Definition 1.1) in order to generate the exchangeable pairs required by [5, Theorem 1.5] (and Lemma B.1), where the event $B_{\eta}^{\square}(p^*)$ is as defined in Equation (30). In the notation of [5, Theorem 1.5], we consider

$$F_{uv}(X, X') := \sum_{\substack{w \in [n] \\ w \notin \{u, v\}}} X_{uw} X_{vw} - \sum_{\substack{w \in [n] \\ w \notin \{u, v\}}} X'_{uw} X'_{vw}.$$

With the help of Lemma B.1, we will now prove Lemma 6.6.

PROOF OF LEMMA 6.6. By Lemma 4.8, we conclude that whenever $p^* \in U_{\beta}$ and $\eta > 0$ is small enough,

$$\mu\left(B_{\eta/2}^{\square}(p^*) \middle| B_{\eta}^{\square}(p^*)^C\right) \leq C(\beta, \eta) \exp(-c(\beta, \eta)n^2).$$

Let X be drawn from the distribution $\mu(\cdot|B_\eta^\square(p^*))$ and let X' be obtained by taking a single step of Glauber dynamics with respect to the measure $\mu(\cdot|B_\eta^\square(p^*))$. Clearly, (X, X') form an exchangeable pair. Let

$$f_{uv}(X) := \mathbb{E}[F_{uv}(X, X')|X].$$

Notice that $|f_{uv}(X)| \leq 1$ almost surely since $|F_{uv}(X, X')| \leq 1$. For n large enough as a function of η , whenever $X \in B_{\eta/2}^\square(p^*)$, the Glauber dynamics over $\mu(\cdot|B_\eta^\square(p^*))$ is exactly equal to the Glauber dynamics w.r.t. μ . Therefore, a simple calculation yields the following for $X \in B_{\eta/2}^\square(p^*)$:

$$(47) \quad f_{uv}(X) = \frac{4p_{uv}(X)}{n-1} - \binom{n}{2}^{-1} \sum_{\substack{w \in [n] \\ w \notin \{u, v\}}} \phi_{uw}(X_{\sim uw}) X_{vw} + \phi_{vw}(X_{\sim vw}) X_{uw}.$$

For any edge e , we consider $\phi_e(X_{\sim e}) = \frac{\exp(\sum_{i=1}^K n^2 \beta_i \Delta_i^e(X))}{1 + \exp(\sum_{i=1}^K n^2 \beta_i \Delta_i^e(X))}$. Note that

$$(48) \quad |\phi_e(X_{\sim e}) - \phi_e(X'_{\sim e})| \leq C(\beta) n^2 \sup_{i \in [K]} |\Delta_i^e(X) - \Delta_i^e(X')|.$$

Now, $X_{\sim e}$ and $X'_{\sim e}$ can differ at most in one edge, by construction. Suppose this edge is h . When $e = h$, then $\Delta_i^e(X) = \Delta_i^e(X')$. Now suppose $e \neq h$. Invoking Equation (42) with $k = |V_i|$, we obtain

$$(49) \quad \begin{aligned} & |\Delta_i^e(X) - \Delta_i^e(X')| \\ &= \left| \int \mathbb{1}(A_e^{kG_i}) \left[\prod_{(a,b) \in E(G)} f^{X_e^+}(x_a, x_b) - \prod_{(a,b) \in E(G)} f^{(X'_e)^+}(x_a, x_b) \right] \prod_{t=1}^k dx_t \right| \\ &\leq \int \mathbb{1}(A_e^{kG_i}) \mathbb{1}(A_h^{kG_i}) \prod_{t=1}^k dx_t \leq \frac{|E(G_i)|^2}{n^3}. \end{aligned}$$

Combining Equations (47), (48) and (49), we conclude that whenever $X \in B_{\eta/2}^\square(p^*)$,

$$(50) \quad |f_{uv}(X) - f_{uv}(X')| \leq \frac{C(\beta)}{n^2}.$$

Observe that whenever the Glauber dynamics does not update an edge of the form (u, w) or (v, w) , $F_{uv}(X, X') = 0$. Let A^{upd} denote the event where $F_{uv}(X, X') \neq 0$. Clearly, $\mathbb{P}(A^{\text{upd}}|X) \leq 4/n$. It is also clear that $|F_{uv}(X, X')| \leq 1$ almost surely. Therefore, we have for any $X \in \mathcal{X}$

$$(51) \quad |f(X)| \leq \mathbb{E}[|F(X, X')| | X] \leq \mathbb{P}(A^{\text{upd}}|X) \leq \frac{4}{n}.$$

Now consider the local variance proxy $\Delta_{uv}(X)$ (where the notation is once again derived from [5, Theorem 1.5]) whenever $X \in B_{\eta/2}^\square(p^*)$,

$$(52) \quad \begin{aligned} \Delta_{uv}(X) &:= \frac{1}{2} \mathbb{E}[(f_{uv}(X) - f_{uv}(X')) F_{uv}(X, X') | X] \\ &\leq \frac{C(\beta)}{n^2} \mathbb{E}[\mathbb{1}(A^{\text{upd}}) | X] \leq \frac{C(\beta)}{n^3}. \end{aligned}$$

Here, we have used Equation (50). Whenever, $X \notin B_{\eta/2}^\square(p^*)$, we will use the crude bound

$$\Delta_{uv}(X) \leq \frac{C}{n^2},$$

obtained by plugging in $|f(X) - f(X')| \leq |f(X)| + |f(X')| \leq 8/n$ (which follows from Equation (51)) into the the definition of Δ_{uv} . Combining these bounds, we get that

$$\Delta_{uv}(X) \leq \frac{C(\beta)}{n^3} \mathbb{1}(X \in B_{\eta/2}^\square(p^*)) + \frac{C(\beta)}{n^2} \mathbb{1}(X \notin B_{\eta/2}^\square(p^*)).$$

By Lemma 4.8,

$$\mu\left(X \notin B_{\eta/2}^\square(p^*) \middle| B_{\eta/2}^\square(p^*)\right) \leq C_{\beta, \eta} \exp(-c(\beta, \eta)n^2).$$

Now, applying Lemma B.1, with $M = 4/n$, and $\theta_0 = C_\beta n^2$ for some large enough C_β and $t = 4\gamma/(n-1)$, we conclude the result. \square

APPENDIX C: DEFERRED PROOFS FOR PATH COUPLING

C.1. Proof of Lemma 4.1. Consider the following coupling between the trajectories X_0, \dots, X_K and Y_0, \dots, Y_K :

1. Generate (X_0, Y_0) from the specified initial distribution.
2. Given X_k, Y_k , we generate X_{k+1}, Y_{k+1} as follows:

$$(53) \quad (X_{k+1}, Y_{k+1}) \sim \begin{cases} Q_{X_k, Y_k} & \text{if } (X_k, Y_k) \in A \\ P(X_k, \cdot) \times P(Y_k, \cdot) & \text{otherwise} \end{cases}$$

Now we consider the distance $d(X_{k+1}, Y_{k+1})$. We have

$$\begin{aligned} \mathbb{E}d(X_{k+1}, Y_{k+1}) &= \mathbb{E}d(X_{k+1}, Y_{k+1}) \mathbb{1}((X_k, Y_k) \in A \times A) \\ &\quad + \mathbb{E}d(X_{k+1}, Y_{k+1}) \mathbb{1}((X_k, Y_k) \in A^c) \\ &\leq \mathbb{E}(1-\gamma)d(X_k, Y_k) \mathbb{1}((X_k, Y_k) \in A) \\ &\quad + \mathbb{E}d(X_{k+1}, Y_{k+1}) \mathbb{1}((X_k, Y_k) \in A^c) \\ &\leq (1-\gamma)\mathbb{E}d(X_k, Y_k) + \mathbb{E}d(X_{k+1}, Y_{k+1}) \mathbb{1}((X_k, Y_k) \in A^c) \\ &\leq (1-\gamma)\mathbb{E}d(X_k, Y_k) + \bar{D}\mathbb{P}((X_k, Y_k) \in A^c) \\ (54) \quad &\leq (1-\gamma)\mathbb{E}d(X_k, Y_k) + \bar{D}p_k. \end{aligned}$$

We conclude the result by unrolling the recursion. \square

C.2. Proof of Lemma 4.10. In both this and the next proof, we let

$$r_{\min}(X) := \inf_{\substack{e \in \binom{[n]}{2} \\ G \in \mathbb{G}_L}} r_G(X, e) \quad \text{and} \quad r_{\max}(X) := \sup_{\substack{e \in \binom{[n]}{2} \\ G \in \mathbb{G}_L}} r_G(X, e).$$

PROOF. We will only prove the coupling for $G(n, p^* + \epsilon)$. The other coupling follows analogously. Let $N = \binom{[n]}{2}$. Let e_1, \dots, e_N be any enumeration of $\binom{[n]}{2}$. Let $X_0 \sim \mu(\cdot | B_\eta^\square(p^*))$ and obtain the sequence X_0, \dots, X_N by updating as follows: Given X_{i-1} , define $(X_i)_e = (X_{i-1})_e$ whenever $e \neq e_i$ and let $(X_i)_{e_i}$ be independently re-sampled from the conditional distribution $(X_{i-1})_{e_i} | (X_{i-1})_{\sim e_i}$. In other words, we obtain X_i by re-sampling the coordinate e_i in X_{i-1} . Clearly, $X_N \sim \mu(\cdot | B_\eta^\square(p^*))$.

Now, consider $Y_0 = 0 \in \mathcal{X}$ almost surely. Let e_1, \dots, e_N be the same as before. We will construct Y_1, \dots, Y_N as follows: Given Y_{i-1} , we construct Y_i such that $(Y_i)_{\sim e_i} = (Y_{i-1})_{\sim e_i}$ and $(Y_i)_{e_i}$ is freshly drawn from $\text{Ber}(p^* + \epsilon)$. It is clear that $Y_N \sim G(n, p^* + \epsilon)$.

By Theorem 4.5, we have that $p^* - \epsilon \leq r_{\min}(X_i) \leq r_{\max}(X_i) \leq p^* + \epsilon$ with probability at-least $1 - C(\eta, \beta, \epsilon) \exp(-c(\eta, \beta, \epsilon)n)$. Recall the definition of ϕ_β from Section 2 and note that $\mathbb{P}((X_i)_{e_i} = 1 | X_{i-1}) \leq \phi_\beta(r_{\max}(X_{i-1})) \leq p^* + \epsilon$. Therefore the fresh draws can be coupled such that $(X_i)_{e_i} \leq (Y_i)_{e_i}$ with probability at least $1 - C(\eta, \beta, \epsilon) \exp(-c(\eta, \beta, \epsilon)n)$. Since $(X_i)_{e_i} = (X_N)_{e_i}$ and $(Y_i)_{e_i} = (Y_N)_{e_i}$, we conclude via a union bound over $i \leq N$ that $X_N \preceq Y_N$ with probability at least $1 - C(\eta, \beta, \epsilon) \exp(-c(\eta, \beta, \epsilon)n)$. This gives the desired coupling by taking $X = X_N$ and $\bar{Y} = Y_N$ in the statement of the lemma. \square

C.3. Proof of Lemma 4.11. We first show that with probability at-least $1 - C(\eta) \exp(-c(\eta)n)$, we must have $p^* - \eta/4 \leq r_{\min}(X_0) \leq r_{\max}(X_0) \leq p^* + \eta/4$.

This can be shown for example by using Lemma 6.1 and simple concentration bounds for the degrees $p_u(X_0)$ and $p_{uv}(X_0)$. The fact that $\delta_{\square}(\tilde{X}_0, p^*)$ is small follows from Theorem 1.3 since $G(n, p^*)$ is also (a very special case of) an exponential random graph. Now, invoking Lemma 4.9, we conclude that with probability at least $1 - TC(\beta, \eta) \exp(-c(\beta, \eta)n)$, we have for every $t \leq T$ that

$$(55) \quad p^* - \frac{\eta}{2} \leq r_{\min}(X_t) \leq r_{\max}(X_t) \leq p^* + \frac{\eta}{2}.$$

Now consider Markov chains Y_0, \dots, Y_T and Z_0, \dots, Z_T where $Y_0 \sim G(n, p^* - \eta/2)$ and $Z_0 \sim G(n, p^* + \eta/2)$. Here, we generate the respective trajectories by Glauber dynamics with respect to $G(n, p^* - \eta/2)$ (resp. $G(n, p^* + \eta/2)$) We couple the trajectories as follows:

1. At step 0, we pick X_0, Y_0, Z_0 such that almost surely

$$Y_0 \preceq X_0 \preceq Z_0.$$

2. At step t , pick the same edge $I_t \sim \text{unif}(\binom{[n]}{2})$ to update for each X_{t-1}, Y_{t-1} and Z_{t-1} .
3. Pick $u_t \sim \text{unif}([0, 1])$ independently of everything else and set

$$(X_t)_{I_t} = \mathbb{1}(u_t \leq \phi_{I_t}((X_t)_{\sim I_t})); \quad (Y_t)_{I_t} = \mathbb{1}(u_t \leq p^* - \frac{\eta}{2}); \quad (Z_t)_{I_t} = \mathbb{1}(u_t \leq p^* + \frac{\eta}{2}).$$

For η small enough, we verify that under the event in Equation (55)

$$p^* - \frac{\eta}{2} \leq \phi_\beta(r_{\max}(X_t)) \leq \phi_{E_t}((X_t)_{\sim E_t}) \leq \phi_\beta(r_{\max}(X_t)) \leq p^* + \frac{\eta}{2}.$$

This implies that $(Y_t)_{I_t} \leq (X_t)_{I_t} \leq (Z_t)_{I_t}$. We conclude that with probability at least $1 - TC(\beta, \eta) \exp(-c(\beta, \eta)n)$ we have $Y_t \preceq X_t \preceq Z_t$. Now, we will apply Theorem 7.2 to obtain

$$(56) \quad \delta_{\square}(\tilde{X}_t, p^*) \leq \max(\delta_{\square}(\tilde{Y}_t, p^*), \delta_{\square}(\tilde{Z}_t, p^*)).$$

Now, since $Y_t \sim G(n, p^* - \eta/2)$ and $Z_t \sim G(n, p^* + \eta/2)$, we have

$$\mathbb{P}\left(\max(\delta_{\square}(\tilde{Y}_t, p^*), \delta_{\square}(\tilde{Z}_t, p^*)) > \eta\right) \leq 1 - C(\eta) \exp(-c(\eta)n^2).$$

Using this in Equation (56), we conclude the statement of the claim. \square

APPENDIX D: PROOF OF COROLLARY 6.3

PROOF. Let $\delta > 0$ and $\eta > 0$. We are given $X \in \mathcal{X}$ such that $X \in B_\eta^\square(p^*)$, i.e., $\delta_{\square}(\tilde{X}, p^*) < \eta$. Consider the set $S(X) := \{u : |p_u(X) - p^*| > 2\eta/\delta\}$. By Lemma 7.1, we note that $|S(X)| \leq \delta n$, so

$$\{S(X) \neq \emptyset\} = \bigcup_{\substack{S \subseteq [n] \\ 1 \leq |S| \leq \delta n}} \{S(X) = S\}.$$

By the union bound,

$$\begin{aligned}
\mu(|S(X)| \neq \emptyset | B_\eta^\square(p^*)) &\leq \sum_{r=1}^{\delta n} \sum_{\substack{S \subseteq [n] \\ |S|=r}} \mu(S(X) = S | B_\eta^\square(p^*)) \\
(57) \quad &= \sum_{r=1}^{\delta n} \binom{n}{r} \mu(S(X) = [r] | B_\eta^\square(p^*)) .
\end{aligned}$$

Here, the second step follows from the permutation invariance of the vertices with respect to the measure μ . In order to evaluate the upper bound in Equation (57), we will consider the measure of the event $\{S(X) = [r]\}$ for $r \leq \delta n$.

Consider the restricted degrees \bar{p}_u with respect to the set $[r]$ and suppose that $\delta^2 < 2\eta$. Let the set $D(p^*, \alpha, S) := \{\mathbf{q}_S : \inf_{u \in S} |\bar{q}_u - p^*| > \alpha\}$ and note that

$$\begin{aligned}
\{S(X) = [r]\} &= \{|p_u(X) - p^*| > 2\eta/\delta, \forall u \in [r]\} \\
&\subseteq \{|\bar{p}_u(X) - p^*| > 2\eta/\delta - \delta, \forall u \in [r]\} \\
(58) \quad &= \bigcup_{\mathbf{q}_{[r]} \in D(p^*, \frac{2\eta}{\delta} - \delta, [r])} \{\bar{p}_u(X) = \bar{q}_u, \forall u \in [r]\} .
\end{aligned}$$

Therefore whenever δ, η satisfy $\epsilon/2 = 2\eta/\delta - \delta$, we have

$$(59) \quad \{S(X) = [r]\} \cap \{X \in B([r], p^*, \eta)\} \subseteq \bigcup_{\mathbf{q}_{[r]} \in D(p^*, \frac{\epsilon}{2}, [r])} A([r], \mathbf{q}_{[r]}, p^*, \eta) .$$

Note that with the above choice of η in the definition of $S(X)$, we can take $\delta < \epsilon/2$ to conclude that for all $u \in S(X)^\complement$, $|p_u(X) - p^*| \leq \epsilon$. We will note a simple result which follows from standard arguments in calculus.

LEMMA D.1. *Suppose $p^* \in U_\beta$ and $q \in [0, 1]$ is such that $|q - p^*| > \epsilon/2$. Then, $L_\beta(q) - L_\beta(p^*) - \zeta_\beta(|q - p^*|) < -C(\beta, \epsilon) < 0$.*

Pick δ to be small enough such that $p^* \in (\delta, 1 - \delta)$. Combining Lemma D.1 with Equation (59) and Theorem 8.9 we conclude that whenever $r \leq n\delta$:

$$\begin{aligned}
\frac{\mu(\{S(X) = r\} \cap B([r], p^*, \eta))}{\mu(B([r], p^*, \eta))} &\leq \sum_{\mathbf{q}_{[r]} \in D(p^*, \frac{\epsilon}{2}, [r])} \frac{\mu(A([r], \mathbf{q}_{[r]}, p^*, \eta))}{\mu(B([r], p^*, \eta))} \\
&\leq |D(p^*, \frac{\epsilon}{2}, [r])| \exp(-nrC(\beta, \epsilon) + O_\beta(nr\eta + r^2 \log(\frac{n}{r}) + r \log n)) \\
&\leq \exp(-2nrC(\beta, \epsilon) + O_\beta(nr\eta + r^2 \log(\frac{n}{r}) + r \log n)) \\
&= \exp\left(-nr\left(2C(\beta, \epsilon) - O_\beta(\eta + \delta \log(\frac{1}{\delta}) + \frac{\log n}{n})\right)\right) \\
(60) \quad &= \exp(-\Omega_{\beta, \epsilon}(nr)) .
\end{aligned}$$

The first step follows from the union bound on Equation (59) and in the second step we used Theorem 8.9 along with Lemma D.1. In the third step, we used that $|D(p^*, \epsilon, [r])| \leq n^r$. In the last step, we have picked η, δ small enough (as functions of ϵ, β) so that $2\eta/\delta - \delta = \epsilon/2$, $\delta < \epsilon/2$, $\eta < \eta_0(\epsilon, \beta)$ and n large enough as a function of ϵ, β such that $O_\beta(\eta + \delta \log(1/\delta) + n^{-1} \log n) < C(\beta, \epsilon)$.

By Lemma 8.6, whenever $r \leq \delta n$, we must have $B_\eta^\square(p^*) = B(\emptyset, p^*, \eta) \subseteq B([r], p^*, \eta) \subseteq B(\emptyset, p^*, \eta + \delta)$. Along with Lemma 4.8, we conclude that whenever η, δ are smaller than some constant $c_\beta > 0$, we have

$$\mu(B_\eta^\square(p^*)) \geq c(\beta, \eta, \delta) \mu(B([r], p^*, \eta)).$$

Therefore whenever n is larger than a constant depending only on β, ϵ and η_0 , we have

$$\begin{aligned} \frac{\mu(\{S(X) = [r]\} \cap B(\emptyset, p^*, \eta))}{\mu(B(\emptyset, p^*, \eta))} &\leq \frac{\mu(\{S(X) = [r]\} \cap B([r], p^*, \eta))}{\mu(B(\emptyset, p^*, \eta))} \\ &\leq C(\beta, \eta, \delta) \frac{\mu(\{S(X) = [r]\} \cap B([r], p^*, \eta))}{\mu(B([r], p^*, \eta))} \\ (61) \quad &\leq \exp(-\Omega_{\beta, \epsilon}(nr)). \end{aligned}$$

Using the fact that $\binom{n}{r} \leq n^r$, it follows using Equations (57) and (61) that

$$\begin{aligned} \mu(|S(X)| \neq \emptyset \mid B_\eta^\square(p^*)) &\leq \sum_{r=1}^{n\delta} n^r \exp(-\Omega_{\beta, \epsilon}(nr)) \\ &\leq \exp(-\Omega_{\beta, \epsilon}(n)), \end{aligned}$$

which proves the corollary. \square

D.1. Proof of Corollary 6.4. Recall the sets $A_u^{k,i} := \{x \in [0, 1]^k : x_i \in [\frac{u-1}{n}, \frac{u}{n}]\}$. Let $\eta > 0$ be arbitrary for now. Suppose that $X \in B_\eta^\square(p^*)$. We have by definition

$$p_{uv}(X) = n^2 \int f^X(x_1, x_3) f^X(x_2, x_3) \mathbb{1}(A_u^{3,1}) \mathbb{1}(A_v^{3,2}) dx_1 dx_2 dx_3.$$

A simple calculation reveals that

$$p_{uv}(X) - p_u(X)p^* = n^2 \int f^X(x_1, x_3) (f^X(x_3, x_2) - p^*) \mathbb{1}(A_u^{3,1}) \mathbb{1}(A_v^{3,2}) dx_1 dx_2 dx_3.$$

Now, define $S_u^+ := \{v : p_{uv}(X) - p^* p_u(X) > \epsilon/2\}$ and $S_u^- := \{v : p_{uv}(X) - p^* p_u(X) < -\epsilon/2\}$. Summing the display above for $v \in S_u^+$, we have

$$\begin{aligned} \frac{|S_u^+| \epsilon}{2n^2} &< \sum_{v \in S_u^+} \int f^X(x_1, x_3) (f^X(x_3, x_2) - p^*) \mathbb{1}(A_u^{3,1}) \mathbb{1}(A_v^{3,2}) dx_1 dx_2 dx_3 \\ &= \int_{x_2 \in A_u^+} f^X(x_1, x_3) (f^X(x_3, x_2) - p^*) \mathbb{1}(A_u^{3,1}) dx_1 dx_2 dx_3 \\ &= \int_{x_1 \in [\frac{u-1}{n}, \frac{u}{n}]} \left[\int_{x_2 \in A_u^+} (f^X(x_3, x_2) - p^*) dx_2 dx_3 \right] dx_1 \\ (62) \quad &\leq \frac{\delta_\square(\tilde{X}, p^*)}{n}, \end{aligned}$$

where $A_u^+ := \bigcup_{v \in S_u^+} [\frac{v-1}{n}, \frac{v}{n}]$ and $C(x_1) := \{x \in [0, 1] : f^X(x_1, x) = 1\}$.

Taking $\eta < \delta\epsilon/2$, we conclude that

$$|S_u^+| \leq \frac{2n\delta_\square(\tilde{X}, p^*)}{\epsilon} \leq \frac{\delta n}{2}$$

and similarly

$$|S_u^-| \leq \frac{2n\delta_{\square}(\tilde{X}, p^*)}{\epsilon} \leq \frac{\delta n}{2}.$$

Therefore, whenever $\delta_{\square}(\tilde{X}, p^*) < \eta < \delta\epsilon/2$, the sets $S_u := S_u^+ \cup S_u^-$ are such that

$$(63) \quad \sup_{u \in [n]} \sup_{v \in S_u^c} |p_{uv}(X) - p^* p_u(X)| \leq \frac{\epsilon}{2}.$$

Now, invoking Corollary 6.3, we conclude that for any $\eta < c(\beta, \epsilon, \delta)$ and whenever n is larger than a constant depending only on β, ϵ, η and δ we have

$$\mu \left(\sup_{u \in [n]} |p_u(X) - p^*| \leq \frac{\epsilon}{2} \middle| B_{\eta}^{\square}(p^*) \right) \geq 1 - \exp(-\Omega_{\beta, \epsilon}(n))$$

and from Equation (63), we have

$$\mu \left(\sup_{u \in [n]} \sup_{v \in S_u^c} |p_{uv}(X) - p^* p_u(X)| \leq \frac{\epsilon}{2} \middle| B_{\eta}^{\square}(p^*) \right) = 1.$$

Combining the two displays above, the statement of the corollary follows. \square