SUB-CRITICAL EXPONENTIAL RANDOM GRAPHS: CONCENTRATION OF MEASURE AND SOME APPLICATIONS

SHIRSHENDU GANGULY AND KYEONGSIK NAM

Abstract. The exponential random graph model (ERGM) is a central object in the study of clustering properties in social networks as well as canonical ensembles in statistical physics. Despite some breakthrough works in the mathematical understanding of ERGM, most notably in [5], through the analysis of a natural Heat-bath Glauber dynamics and in [14, 20, 21], via a large deviation theoretic perspective, several basic questions have remained unanswered owing to the lack of exact solvability unlike the much studied Curie-Weiss model (Ising model on the complete graph). In this paper, we establish a series of new concentration of measure results for the ERGM throughout the entire sub-critical phase, including a Poincaré inequality, Gaussian concentration for Lipschitz functions, and a central limit theorem. In addition, a new proof of a quantitative bound on the $W_1$–Wasserstein distance to Erdős-Rényi graphs, previously obtained in [45], is also presented. The arguments rely on translating temporal mixing properties of Glauber dynamics to static spatial mixing properties of the equilibrium measure and have the potential of being useful in proving similar functional inequalities for other Gibbsian systems beyond the perturbative regime.

CONTENTS

1. Introduction .............................. 1
2. Statements of the results ................. 8
3. Main ideas of the proofs and organization of the article .................. 11
4. ERGM in the high temperature regime .... 13
5. Gaussian concentration: Proof of Theorem 1 .................. 20
6. Central limit theorem: Proof of Theorem 2 .................. 23
7. Bounding $W_1$–Wasserstein distance: Proof of Theorem 3 ............ 29
References .................................. 33

1. INTRODUCTION

A central object in the study of statistical models on networks, is the notion of a Gibbs measure on graphs. In the most general setting, the probability of a graph $G$ on $n$ vertices, thought of naturally as $x \in \{0,1\}^{n(n-1)/2}$ is proportional to $e^{\beta f(x)}$ where $f(\cdot)$ is the Hamiltonian and $\beta$ is a parameter classically referred to as the inverse temperature.

A particularly important subclass of such measures, capturing clustering properties, is obtained when the Hamiltonian is given by counts of subgraphs of interest, such as triangles. This is termed
in the literature as the *Exponential Random Graph model* (ERGM). Thus more precisely, for \( x \in \{0, 1\}^{n(n-1)/2} \), where the configuration space is the set of all graphs on the vertex set \( \{1, \ldots, n\} \), defining \( N_G(x) \) to be the number of labeled subgraphs \( G \) in \( x \), given a vector \( \beta = (\beta_1, \ldots, \beta_s) \), the ERGM Gibbs measure is defined as

\[
\pi(x) \sim \exp \left( \sum_{i=1}^{s} \beta_i \frac{N_{G_i}(x)}{n^{|V(G_i)|-2}} \right),
\]

(see Section 1.2 for details) and hence is a version of the well known Erdős-Rényi graphs, obtained by *tilting* according to the subgraph counting Hamiltonian. Being mostly used for modeling relational networks in sociology, there is a significant amount of rigorous and non-rigorous literature on ERGM, see e.g. [24, 28, 42, 43] for some specific cases, while [13] verified a mean-field approximation for some values of \( \beta \), where the ERGM behaves qualitatively like an Erdős-Rényi graph. There is also a series of works studying constrained ERGM models, as an important example of microcanonical ensembles in statistical physics (see e.g. [30] and the references therein). A characteristic property of the ERGM is the well known mean field behavior, which informally means that it can be approximated in an information theoretic sense by product measures. Unfortunately, contrary to classical spin systems and lattice gases, a detailed analysis of general ERGM has been out of reach so far. Thus, while a lot of refined results on fluctuation theory and concentration properties, have been established over the years, for the exactly solvable Curie-Weiss model, (Ising model on the complete graph), corresponding questions for the ERGM remain largely open.

However a significant breakthrough was made in [5], who studied a natural Heat-bath Glauber dynamics on \( G_n \) with ferromagnetic ERGM as the invariant measure, and established precise estimates on convergence to equilibrium as well as closeness to an Erdős-Rényi measure of appropriate density as a function of \( \beta \).

Soon after, in a landmark development Chatterjee and Diaconis [14] introduced a large deviation theoretic approach to the study of ERGM. Among many things, one of the key achievements of this work, is a variational formula for the free energy of ERGM, using the large deviation theory for Erdős-Rényi graphs, developed in prior work [15] relying on the theory of graph limits developed by Lovász and coauthors [35]. Analyzing the formula, it was also established that ERGM for certain values of \( \beta \) behaves qualitatively like an Erdős-Rényi graph (in an entropy theoretic sense) in the thermodynamic limit. More recently, a more refined result was established by Eldan [20], who obtained a representation of the ERGM as a low entropy mixture of product measures, using the framework of nonlinear large deviation theory developed in the same paper, extending the theory put forward in [12].

Recently, the results in [5] have been extended to what are known as vertex weighted exponential random graphs in [17]. We refer to [21, 44, 47, 51] for more results on ERGM and [11] for a beautiful exposition of the recent developments around the general theory of large deviations for random graphs.

In spite of the above developments, the understanding of the ERGM was still not refined enough to treat important and delicate questions about concentration of measure properties and central limit theorems with the exception of the perturbative, very high temperature regime, popularly
known as *Dobrushin’s uniqueness (DU) regime* where [46] showed that the ERGM behaves qualitatively like a product measure, satisfying the Log-Sobolev inequality (LSI) and as a consequence, strong concentration properties.

However, going beyond perturbative ideas, while for classical spin systems and related percolation models, there has been significant progress in understanding probabilistic aspects in the various phases of temperature, analogous questions for non-exactly solvable mean-field models like the ERGM, about spectral gap, Log-Sobolev inequality, concentration of measure, and central limit theorems are largely open.

To elaborate on a concrete difficulty one faces, let us, for example, consider the concentration phenomenon for the Ising model on the lattice, which can be obtained by a martingale argument or coercivity properties such as Poincaré and Log-Sobolev inequality. The validity of these conditions require a spatial mixing property which has been verified in some cases throughout the entire high temperature phase, on the lattice (see [2] for the related results for the Ising model on general graphs). In fact, it is known that a certain mixing condition, well known as the *Dobrushin-Shlosman mixing condition*, is equivalent to the uniform boundedness of the Log-Sobolev constant (see [36, 37, 50, 49] for details). However, while spatial mixing properties have been an object of intense study for finite range spin systems, such properties cannot be expected for the ERGM owing to the natural symmetry, exchangeability and the mean field nature of the model.

In this paper, we establish a series of new results for the ERGM which hold throughout the *whole high temperature regime*. The main result establishes Gaussian concentration for Lipschitz functions, and as applications we prove a central limit theorem (CLT) for a partial number of edges as well as a quantitative result about how close the high temperature ERGM is to an Erdős-Rényi graph. We also establish a sharp Poincaré inequality. Most of the results in the paper are the first of their kind beyond the perturbative regime for the ERGM. We also include a discussion about potential applications of some of the results in this paper to problems in statistics about estimability of parameters in an ERGM, which along with the CLT result, can lead to progress in the analysis of natural hypothesis testing problems on networks.

The key ingredient we rely on is a temporal mixing result along the Heat-bath Glauber dynamics derived from [5]. Thus the main theme in this paper is how such estimates can be translated into mixing properties of the equilibrium measure which we expect would be of general interest particularly in analyzing other related Gibbs measures.

We now move on to the precise forms of the main results and start by developing the necessary notation.

1.1. **Configuration space and notations.** For any graph $G$, we will use $V(G)$ and $E(G)$ to denote the vertex and edge sets, respectively. Let $\mathcal{G}_n$ be the set of all graphs with vertex set $[n] := \{1, 2, \ldots, n\}$. In particular, let $K_n$ be a complete graph with a vertex set $[n]$. Adopting a widely used notation, an element in $\mathcal{G}_n$ will be denoted by $x = (x_e)_{e \in E(K_n)}$ with $x_e = 1$ if the (undirected) edge $e = (i,j)$ is present and $x_e = 0$ otherwise, i.e., we will identify $\mathcal{G}_n$ with the hypercube $\{0,1\}^M$, where $M := \frac{n(n-1)}{2}$.

For $x \in \mathcal{G}_n$, let $E(x)$ be the set of edges $e$ with $x_e = 1$. Define $\bar{x}_e = (x_f)_{f \neq e}$ to be the collection of all coordinates of $x$ except for the edge $e$. Also, denote $x_{e^+}$ and $x_{e^-}$ to be configurations whose
edge sets are $E(x) \cup e$ and $E(x) \setminus e$, respectively. Further, let $x_{ef,+}$ be a configuration whose edge set is $E(x) \cup e \cup f$. Finally, for any function $\varphi : \mathcal{G}_n \to \mathbb{R}$, let
\begin{align}
\partial_e \varphi(x) &= \varphi(x_{e+}) - \varphi(x_{e-}), \\
\partial_{ef} \varphi(x) &= \partial_e \varphi(x_{f+}) - \partial_e \varphi(x_{f-})
\end{align}

denote the discrete first and second derivatives.

We now define a natural partial ordering and metric on $\mathcal{G}_n$. We say $x \leq y$, if and only if $x_e \leq y_e$ for all edges $e$. For $x, y \in \mathcal{G}_n$, define $x \wedge := (x_e \wedge y_e)_{e \in E(K_n)}$ with $x_e \wedge y_e = \min\{x_e, y_e\}$ and $x \vee := (x_e \vee y_e)_{e \in E(K_n)}$ with $x_e \vee y_e = \max\{x_e, y_e\}$. We will use $d_H$ to denote the Hamming distance on $\mathcal{G}_n$:
\begin{equation}
d_H(x, y) := \sum_{i=1}^{M} |x_i - y_i|.
\end{equation}

Next, we discuss the crucial notion of the subgraph counting number. For any graph $G$, let $N_G(x)$ be the number of copies of the graph $G$ in the configuration $x$, multiplied by the number of automorphisms of $G$. More precisely:
\begin{equation}
N_G(x) = \sum_\psi \prod_{(v, w) \in E(G)} x(\psi(v), \psi(w)),
\end{equation}

where the sum is over all injective function $\psi : V(G) \to [n]$. For instance, if $G$ is a triangle, then $N_G(x)$ is the number of labeled triangles in $x$. For our purposes, we would also need slight variants of the above definition. For a configuration $x$ and an edge $e = (i_1, i_2)$, let $\hat{x} := x_{e,+}$ and define,
\begin{equation}
N_G(x, e) := \sum_\psi \prod_{(v, w) \in E(G)} \hat{x}(\psi(v), \psi(w)),
\end{equation}

where the sum is over all injective function $\psi : V(G) \to [n]$ satisfying $i_1, i_2 \in \psi(V(G))$ and $(\psi^{-1}(i_1), \psi^{-1}(i_2)) \in E(G)$, i.e., $N_G(x, e)$ counts all embeddings of $G$ into $x_{e,+}$, where some edge of $G$ maps to $e$. Likewise, for a configuration $x$ and two distinct edges $e = (i_1, i_2)$ and $f = (i_3, i_4)$, letting $\hat{x} = x_{ef,+}$, define,
\begin{equation}
N_G(x, e, f) := \sum_\psi \prod_{(v, w) \in E(G)} \hat{x}(\psi(v), \psi(w)),
\end{equation}

where the sum is again over all injective function $\psi : V(G) \to [n]$ satisfying $i_1, \ldots, i_4 \in \psi(V(G))$ and $(\psi^{-1}(i_1), \psi^{-1}(i_2), \psi^{-1}(i_3), \psi^{-1}(i_4)) \in E(G)$. Note that above, $e$ and $f$ are allowed to share a vertex. For $e = f$, set $N_G(x, e, f) := 0$. The latter convention is rather natural because of the following reason. For the Hamiltonian $H$, defined shortly in (8), which will be considered throughout the paper, for $e \neq f$, $\partial_{ef} H(x)$ has a natural expression in terms of $N_G(x, e, f)$ (see e.g., (36)). Since by definition $\partial_{ee} H(x) = 0$, we adopt the convention of defining $\hat{N}_G(x, e, e) := 0$ so that (36) holds for all $e, f$.

Letting $\mathcal{P}(\mathcal{G}_n)$ be the collection of probability measures on $\mathcal{G}_n$, for any $\mu \in \mathcal{P}(\mathcal{G}_n)$, $\mu_{e}(\cdot|x)$ denotes the conditional distribution of the edge $e$ given $\bar{x}_e$. Finally, for any $\mu, \nu \in \mathcal{P}(\mathcal{G}_n)$, let
\begin{equation}
d_{TV}(\mu, \nu) := \sup_{\Lambda \subseteq \{0, 1\}^M} |\mu(A) - \nu(A)|,
\end{equation}

denote the total variation distance between $\mu$ and $\nu$, respectively.
1.2. The Hamiltonian and the Gibbs measure. We now arrive at the definition of the Gibbs measure on $\mathcal{G}_n$, popularly known as the exponential random graph model (ERGM), which is the central object of study in this article. Fix $s$ many graphs $G_1, \cdots, G_s$ with $|V_i| := |V(G_i)|$ and $|E_i| := |E(G_i)|$. Let $a$ be a positive integer satisfying $|V_i| \leq a$ for $i = 1, \cdots, s$. We define the Hamiltonian $H$ on $\mathcal{G}_n$ by
\[ H(x) = \sum_{i=1}^{s} \beta_i \frac{N_{G_i}(x)}{|V_i| - 2}, \quad (8) \]
where $\beta_i$ are certain parameters which will be encoded by the vector $\beta$. Note that the subgraph count $N_{G_i}(x)$ is of order $n|V_i|$; thus, the normalization $n|V_i| - 2$ ensures that the Hamiltonian is of order $n^2$, which turns out to be the right scaling for the model. Finally, the ERGM($\beta$) (often to be called as just the ERGM) is the Gibbs measure $\pi \in \mathcal{P}(\mathcal{G}_n)$ defined by
\[ \pi(x) = \frac{1}{Z_n(\beta)} \exp \left( \sum_{i=1}^{s} \beta_i \frac{N_{G_i}(x)}{|V_i| - 2} \right), \quad (9) \]
where $Z_n(\beta)$ is the normalizing constant and the quantity $f_n(\beta) := \frac{\log Z_n(\beta)}{n^2}$, will be called the normalized free energy. Throughout the paper, we only consider the case when $\beta_i > 0$ for $i = 1, \cdots, s$ so that the Gibbs measure (9) is monotone and ferromagnetic.

1.3. Glauber dynamics. There is a natural (discrete time) Heat-bath Glauber dynamics (GD) associated with the ERGM (9), which is defined as follows. Given the current state $x$, an edge $e$ is uniformly chosen and resampled according to the conditional distribution of $x_e$ given $\bar{x}_e$. It is an easy calculation to verify that given $e$ is chosen to be updated, one has the following transition probabilities:
\[ P(x, x_{e^+}) = \frac{\exp(\partial_e H(x))}{1 + \exp(\partial_e H(x))}; P(x, x_{e^-}) = \frac{1}{1 + \exp(\partial_e H(x))}, \quad (10) \]
and as is well known for Glauber dynamics, $\pi$ is reversible with respect to the transition kernel. For $x \in \mathcal{G}_n$ and edge $e$, let $x^e$ be the configuration $x$ with edge $e$ flipped (i.e. open $\leftrightarrow$ closed), and define $c(x, e) = P(x, x^e)$ according to (10). Then, the generator of Glauber dynamics $\mathcal{L}$ is defined by
\[ \mathcal{L} f(x) = \frac{1}{M} \sum_{e} c(x, e)(f(x^e) - f(x)), \quad (11) \]
and the Dirichlet form $\mathcal{E}$ is given by
\[ \mathcal{E}(f, g) := \frac{1}{M} \sum_{x \in \mathcal{G}_n} \sum_{e} c(x, e)(f(x^e) - f(x))(g(x^e) - g(x))\pi(x). \quad (12) \]
Also, for $t \geq 0$, the semigroup generated by $\mathcal{L}$ will be denoted by
\[ P_t := e^{t\mathcal{L}}. \]

We next introduce the grand coupling, which provides a natural coupling between the GD, $\{(Z^x(t))_{x \in \mathcal{G}_n}\}_{t \geq 0}$, starting from all initial configurations $x$ and is often an useful tool to analyze Markov chains.
Because of its appearance throughout the paper, it would be convenient to first let

\[ \Phi(x) := \frac{e^x}{1 + e^x}. \] (13)

Then, for \( I \), a uniformly chosen edge, and \( U \) a uniform random variable on \([0, 1]\) independent of \( I \), for \( x \in \mathbb{G}_n \), define,

\[ S^x := \begin{cases} 1 & \text{if } 0 < U \leq \Phi(\partial_I H(x)), \\ 0 & \text{if } \Phi(\partial_I H(x)) < U \leq 1, \end{cases} \] (14)

and proceed to define the next state \( Z^x(1) \), as

\[ Z^x(1) = \begin{cases} x_e & \text{if } e \neq I, \\ S^x & \text{if } e = I. \end{cases} \]

The above described one step of the grand coupling can be repeated in a Markovian fashion to obtain the full grand coupling. We will also denote the full and empty configurations by \(+\) and \(−\) respectively, and \( Z^+(t) \) and \( Z^-(t) \) to denote the GD started from them. A particularly useful property of the grand coupling is that it is monotone, i.e., the natural partial order on the configuration space defined earlier is preserved in time under the coupling and hence we will interchangeably refer to this as the monotone coupling. Also for any measure \( \mu \in \mathcal{P}(\mathbb{G}_n) \), we will denote the distribution of the GD at time \( t \) starting from \( \mu \) by \( \mu_t \).

Finally we denote the \( \epsilon \)-mixing time by

\[ t_{\text{mix}}(\epsilon) := \inf \{ t : d_{\text{TV}}(Z^x(t), \pi) \leq \epsilon \ \forall \ x \in \mathbb{G}_n \}. \]

1.4. High and low temperature phases. To motivate the objects of interest in this paper, distinguish the high and low temperature behaviors and put our results in context, it will be useful to review briefly the prior advances in the study of ERGM. Broadly the two main breakthroughs in this field rely on two different perspectives, large deviation theoretic and the study of GD. Using the former approach, in a seminal work, the thermodynamic behavior of ERGM was studied by Chatterjee and Diaconis [14] who obtained, a precise variational formula for the free energy of ERGM, \( f_n(\beta) \), via the theory of graphons which are measurable and symmetric functions \( f : [0, 1]^2 \to [0, 1] \). More formally, generalizing the notion of Gibbs measures on graphs, denoting the space of all graphons modulo composition by measure preserving transformation of \([0, 1]\) to itself, by \( \tilde{W} \), for a function \( T : \tilde{W} \to \mathbb{R} \), one can consider the probability distribution \( p_n \) on \( \tilde{G}_n \) defined by

\[ p_n(G) = \frac{1}{Z_n} e^{n^2 T(\tilde{G})}, \]

where \( \tilde{G} \) is a canonical embedding of \( G \) into \( \tilde{W} \), and \( Z_n \) is the corresponding normalizing constant. Given the above setting, using a previous result of Chatterjee and Varadhan [15] who established a large deviation principle on \( \tilde{W} \), for the sequence of measures induced by the Erdős-Rényi \( G(n, p) \) graph for fixed \( p > 0 \) (dense case) as \( n \to \infty \), Chatterjee and Diaconis [14] showed that,

\[ \lim_{n \to \infty} \frac{\log Z_n}{n^2} = \sup_{\tilde{h} \in \tilde{W}} \left( T(\tilde{h}) - \frac{1}{2} I(\tilde{h}) \right), \] (15)

where for any \( x \in [0, 1] \), we let

\[ I(x) := -[x \log x + (1 - x) \log(1 - x)] \]
to be the binary-entropy functional, while for any graphon $W$, $I(W)$ is simply $\int_{[0,1]^2} I(W(x,y)) dxdy$.

**Replica-Symmetry:** In particular, they proved that the ferromagnetic case $\beta_1, \ldots, \beta_s > 0$ falls in the replica symmetric regime i.e., the maximizers in (15) are given by the constant functions, i.e.,

$$\lim_{n \to \infty} \frac{1}{n^2} \log Z_n(\beta) = \sup_{u \in [0,1]} \left( \sum_{i=1}^{s} \beta_i u|E_i| - \frac{1}{2} I(u) \right).$$  \hspace{1cm} (16)

Informally, the above implies that if $u_1, \ldots, u_k \in [0,1]$ are the maximizers in (16), then the ERGM (9) behaves like a mixture of the Erdős-Rényi graphs $G(n,u_i)$’s, in an asymptotic sense. This was made precise in the more recent work by Eldan [20], and Eldan and Gross [21].

To understand the solutions of (16), recalling the function $\Phi$ from (13), for $p > 0$, define two functions $\Psi_\beta$ and $\varphi_\beta$ by

$$\Psi_\beta(p) := \sum_{i=1}^{s} 2\beta_i |E_i| p|E_i|^{\frac{1}{2}}, \quad \varphi_\beta(p) := \Phi(\Psi_\beta(p)) = \frac{\exp(\Psi_\beta(p))}{1 + \exp(\Psi_\beta(p))};$$  \hspace{1cm} (17)

It is not difficult to check that both the above functions are increasing in $p$. Note that by computing the gradient of the RHS in (16), all the maximizers of the same, satisfy

$$\varphi_\beta(p) = p.$$  \hspace{1cm} (18)

However, while the solutions of the RHS in (16) represent the constant graphons which account for the dominant contribution to the free energy, there might be several solutions to (18), representing more local maxima. It is not difficult to see that the second order condition associated to a $p$ corresponding to a local maxima for (16) is

$$\varphi'_\beta(p^*) < 1.$$  \hspace{1cm} (19)

In fact such states have a natural interpretation in terms of the GD. This perspective was adopted in another major advancement in this field in [5] where the authors pursued an alternate approach of understanding ERGM through studying convergence to equilibrium for the GD. Following [5], we say that $\beta$ belongs to the high temperature phase or is subcritical if $\varphi_\beta(p) = p$ has a unique solution $p^*$ and, which in addition satisfies $\varphi'_\beta(p^*) < 1$. Thus in the subcritical phase the mixture of Erdős-Rényi graphs $G(n,u_i)$ degenerates to a single Erdős-Rényi graph $G(n,p^*)$. Whereas, $\beta$ is said to be in the low temperature phase or is supercritical, if $\varphi_\beta(p) = p$ has at least two solutions $p^*$ with $\varphi'_\beta(p^*) < 1$. When $\beta$ is neither in the high nor low temperature phase, is called the critical temperature phase. It was shown in [5] that GD is rapidly mixing if $\beta$ is subcritical, and is slowly mixing if it is supercritical instead. More generally, it is not difficult to, at least heuristically, argue that any local maxima, i.e., $p$ satisfying (18) and (19) represents a local metastable state in the following sense: starting from an Erdős-Rényi graph $G(n,p)$, under the GD evolution, the edge and other subgraph densities stay close to that of the starting state for an exponential in $n$ amount of time. A related version of such a result appears in the low temperature analysis in [5].
2. Statements of the results

Even though the above results establish in a certain weak sense, that ERGM in the high temperature phase behaves like an Erdős-Rényi graph $G(n,p^*)$, several problems remain open, in particular pertaining to how this approximation can be made quantitative. We list below a few of them:

1. Does the Glauber dynamics satisfy functional inequalities like Poincaré and Log-Sobolev inequalities?
2. What kind of concentration of measure does the ERGM exhibit?
3. Do natural observables like the number of edges in an ERGM satisfy a central limit theorem?

Answers to such questions have several potential applications including in testing of hypothesis problems in the statistical study of networks and not surprisingly, this has led to a significant body of work over the recent years. Given the above preparation, we now present our main results and compare them to the progress made in the aforementioned, related works. Throughout the sequel, we assume that $\beta$ is in the sub-critical (high temperature) phase. Furthermore, since we will have to deal with different measures, such as the equilibrium measure $\pi$, the measure induced by the Glauber dynamics, various couplings in the definitions of Wasserstein distances and so on, we will use the same notation $\mathbb{P}$ to denote the underlying measure and the corresponding expectation by $E$. The exact object will be clear from context and will not lead to any confusion. Finally a random graph of size $n$ will be denoted by $X = (X_e)_{e \in E(K_n)}$, where $X_e$ is the indicator that the edge $e$ is present.

**Spectral-gap.** We will denote the spectral gap of the generator $\mathcal{L}$ defined in (11) by $\gamma$. The following variational characterization is well known:

$$\gamma = \inf_{f} \mathcal{E}(f,f), \text{ where } f : G_n \to \mathbb{R} \text{ satisfies } E_\pi(f) = 0 \text{ and } E_\pi(f^2) = 1. \quad (20)$$

It is proved in [5], that the mixing time of GD in the high temperature phase is $\Theta(n^2 \log n)$, which in turn implies that the relaxation time i.e., $\frac{1}{\gamma}$, is $O(n^2 \log n)$. However it turns out that as a relatively straightforward consequence of already known results, a sharp estimate of the relaxation time throughout the entire high temperature regime follows. This is the first result we record.

**Theorem 2.1.** The spectral gap $\gamma$ of ERGM is $\Theta(n^{-2})$. Thus there exists a constant $C > 0$ such that for sufficiently large $n$ and any $f \in L^2(\pi)$,

$$\text{Var}(f) \leq Cn^2 \mathcal{E}(f,f). \quad (21)$$

By standard theory of reversible Markov chains [34, Chapter 12], the above implies that the variance of $P_t f$ decays exponentially fast along the GD, i.e., there exists $c > 0$ such that for sufficiently large $n$ and any $t > 0$,

$$\text{Var}(P_t f) \leq \left(1 - \frac{c}{n^2}\right)^t \text{Var}(f).$$

A similar bound holds for the Dirichlet form $\mathcal{E}(P_t f, P_t f)$ as well,

$$\mathcal{E}(P_t f, P_t f) \leq \left(1 - \frac{c}{n^2}\right)^t \mathcal{E}(f,f).$$

**Concentration of measure.** Concentration of measure properties of ERGM are the focus of this article. Often in spin systems, for very high temperature values, the model falls in DU regime
(see Section 4.1 for definition), where certain classical perturbation arguments can be used to prove Gaussian concentration for Lipschitz functions. While for the ERGM, similar arguments were presented in [46], going beyond the DU regime has resisted rigorous mathematical analysis so far. Before making precise statements, we first define the notion of a Lipschitz function on $G_n$: for an $M$-dimensional vector $v = (v_1, \cdots, v_M)$ with $v_i \geq 0$, and $f : G_n \mapsto \mathbb{R}$ we say that $f \in \text{Lip}(v)$ if
\[
|f(x) - f(y)| \leq \sum_{i=1}^{M} v_i \mathbb{1}\{x_i \neq y_i\}. \tag{22}
\]

A metric measure space $(X, d, \mu)$ is said to admit an exponential concentration if for some $u > 0$,
\[
\sup_{\|f\|_{\text{Lip}} \leq 1} \int_X e^{uf} d\mu < \infty.
\]
It is a classical fact that when $X$ is a Riemannian manifold with metric $d$ and induced volume measure $\mu$, the existence of a positive spectral gap for the Laplacian implies the exponential concentration of $(X, d, \mu)$. We refer to the monographs [6, 32] for more details.

The discrete analog of the above fact was obtained by Ledoux [33]. More formally, let $X$ be a finite (or countable) set, and $P : X \times X \to \mathbb{R}$ be a Markov kernel. Also assume that a probability measure $\mu$ is reversible with respect to $P$. Then, we define the following notion of the Lipschitz constant of a function $f : X \to \mathbb{R}$:
\[
\|f\|_{\text{Lip}}^2 := \sup_{x \in X} \sum_{y \in X} |f(x) - f(y)|^2 P(x, y), \tag{23}
\]
and then define the canonical distance on $X$ by
\[
d(x, y) := \sup_{\|f\|_{\text{Lip}} \leq 1, \int f d\mu = 0} [f(x) - f(y)].
\]
It is proved in [33, Theorem 2.2] that if $\lambda_1$ is a spectral gap of the Markov chain, then
\[
\sup_{\|f\|_{\text{Lip}} \leq 1} \int_X e^{\sqrt{\lambda_1/2}f} d\mu < 3.
\]
In particular, $(X, d, \mu)$ admits an exponential concentration if $\lambda_1 > 0$. Applying this fact in conjunction with the already established Poincaré inequality (21) for the ERGM, allows one to conclude that there exists a constant $C > 0$ such that for any $f$ with $\|f\|_{\text{Lip}} \leq 1$ (with respect to the Lipschitz norm (23)),
\[
\mathbb{P}(|f(X) - \mathbb{E}f(X)| > t) \leq 3e^{-Ct/n}.
\]

The next theorem is the main result of this article which improves the above to establish the expected Gaussian concentration result throughout the high temperature phase.

**Theorem 1.** There exists a constant $c > 0$ such that for sufficiently large $n$, for any $f \in \text{Lip}(v)$ and $t \geq 0$,
\[
\mathbb{P}(|f(X) - \mathbb{E}f(X)| > t) \leq 2e^{-c t^2/\|v\|_1 \|v\|_\infty}. \tag{24}
\]
Central limit theorem. Going beyond concentration of measure, establishing central limit theorems (CLT) has been a fundamental problem in the study of classical Gibbsian systems. For instance, CLT for the magnetization has been obtained for the subcritical (finite range) Ising model, using exponential decay of correlations (see \cite{22,27,40,41} for more details). A similar result holds for the exactly solvable mean-field Ising model (Curie-Weiss) for all sub-critical $\beta < 1$ (see \cite{23} for more details).

However, unlike classical spin systems, as has been alluded to before, there are several barriers to proving a CLT type result in the ERGM setting. The only result we are aware of is for the ERGM with the Hamiltonian being the two-star counting function (two-star model) where \cite{39} established a full CLT heavily exploiting a special Ising-type interaction the model admits.

Our next result makes partial progress in this front. In particular, we prove a CLT throughout the high temperature regime for the number of open bits restricted to a sub-linear (in $n$) number of coordinates. To the best of the authors’ knowledge, this is the first CLT of any kind in this general setting. While the full CLT for the number of edges is still open, refined versions of the methods introduced in this paper might be useful in resolutions of the same and other related questions.

**Theorem 2.** For any sequence of positive integers $m$ satisfying $m = o(n)$ and $m \to \infty$ as $n \to \infty$ we have the following. Consider any set of $m$ different edges $i_1, \ldots, i_m \in E(K_n)$ that do not share a vertex. Then, the following central limit theorem for the normalized number of open edges among $i_1, \ldots, i_m$ holds:

$$\frac{X_{i_1} + \cdots + X_{i_m} - E[X_{i_1} + \cdots + X_{i_m}]}{\sqrt{\text{Var}(X_{i_1} + \cdots + X_{i_m})}} \xrightarrow{d} N(0,1),$$

where $\xrightarrow{d}$ denotes weak convergence.

Bounding the $W_1$–Wasserstein distance. Our final result is another consequence of the Gaussian concentration result Theorem 1. Namely, we provide an alternate proof of a quantitative estimate on the closeness between a high temperature ERGM and an Erdős–Rényi graph obtained in \cite{45}. It is not hard to see from (17), that in the multiparameter case $s \geq 2$, the set of $\beta = (\beta_1, \ldots, \beta_s)$’s, in the high temperature phase, sharing the same $p^*$, form a $(s-1)$ dimensional hyperplane solving

$$p^* = \frac{\exp(\sum_{i=1}^s 2\beta_i |E_i|(p^*)^{E_i|-1}}{1 + \exp(\sum_{i=1}^s 2\beta_i |E_i|(p^*)^{E_i|-1}}.$$  

Thus from the discussion so far in this article, for all these $\beta$, the ERGM behaves qualitatively like the Erdős–Rényi graph $G(n, p^*)$ in various senses. A particular way to quantify the same is through the notion of contiguity\(^1\), leading one to speculate whether any ERGM($\beta$) satisfying (26), is contiguous with the Erdős–Rényi graph $G(n, p^*)$. While such results have been indeed recently proved for the multi-parameter Ising model in \cite{26}, for the ERGM, in the special case

\(^1\)Given two sequences of measures $\{p_n\}, \{q_n\}$ where $p_n$ and $q_n$ are defined on the same space, one says that the former is contiguous with respect to the latter, if for any sequence of events $A_n$, $q_n(A_n) \to 0$ implies $p_n(A_n) \to 0$. 

This is the first result of its kind, and we expect that the method could be useful in other related settings. We next move on to a key application of the above.
where the Hamiltonian $H(x)$ depends only on the number of edges and the two-star count, [39] in fact constructed a consistent estimator of $\beta$ for all values of the latter, proving that the two measures are in fact not contiguous. While a general result refuting contiguity is still missing, currently in the sub-critical case, we present a quantitative bound in terms of the $W_1$–Wasserstein distance between $\text{ERGM}(\beta)$ and $G(n,p^*)$, where for $\mu, \nu$ in $\mathcal{P}(G_n)$:

$$W_1(\mu, \nu) = \inf_{\Gamma} \sum_{e \in E(K_n)} \mathbb{P}_\Gamma(X_e \neq Y_e),$$

where the infimum is obtained over all couplings $\Gamma$ of $\mu$ and $\nu$.

**Theorem 3.** There exists a constant $C > 0$, such that for sufficiently large $n$,

$$W_1(\pi, \nu) \leq C n^{3/2} \sqrt{\log n},$$

(27)

where $\pi$ is the $\text{ERGM}(\beta)$ distribution, and $\nu \in \mathcal{P}(G_n)$ is the Erdős-Rényi distribution $G(n,p^*)$.

The above result is of relevance in the study of estimability properties of the ERGM and a related discussion is presented later in Section 7.1.

**Remark 2.2.** A slightly stronger statement than (27), without the logarithmic term $\sqrt{\log n}$, is obtained in [45]. It is shown that for any sub-critical $\text{ERGM}(\beta)$, there exists a constant $C > 0$ such that for sufficiently large $n$ and any Lipschitz function $f$,

$$|\mathbb{E}f(X) - \mathbb{E}f(Y)| \leq C \|f\|_{Lip} \cdot n^{3/2},$$

(28)

where, the Lipschitz constant $\|f\|_{Lip}$ is defined as an infimum of values $c > 0$ such that

$$|f(x) - f(y)| \leq c \sum_{i=1}^{M} |x_i - y_i|$$

for any $x, y \in G_n$. The key step in [45] to obtain (28) was to use the Stein’s method for CLT and obtain a solution $h$ to Stein equation

$$\mathcal{L}h(x) = f(x) - \mathbb{E}f(X)$$

(recall that $\mathcal{L}$ is a generator associated with $\text{ERGM}(\beta)$).

3. **Main ideas of the proofs and organization of the article**

We discuss the key ideas in the proofs of the theorems stated above.

3.1. **Poincaré inequality.** The proof of this, in fact, follows in a relatively straightforward fashion from known results. Nonetheless, we include it for completeness. To understand the spectral gap, we rely on its relation to temporal mixing properties of the GD, which was the subject of study in [5]. It is a classical fact (see e.g., [34]) that a Markov chain exhibiting a strict contraction property (described below), has a positive spectral gap. In particular, this is verifiable if the temperature $\beta$ lies in DU regime. In fact, if the $\beta_i$'s are small enough so that $\Psi(1) < 4$ (see (17)), then there exists a constant $\delta > 0$, such that for all $x, y \in G_n$ some coupling of $Z^x(1)$ and $Z^y(1)$,

$$\mathbb{E}d_H(Z^x(1), Z^y(1)) < \left(1 - \frac{\delta}{n^2}\right)d_H(x, y)$$

(29)
(more explanations are provided in Section 4.1). This immediately implies that the relaxation time is $O(n^2)$ in DU regime.

However, (29) does not hold throughout the high temperature phase. Nonetheless, it is proved in [5] that there exists a constant $c > 0$ such that something quite similar does hold. Namely, under the monotone coupling $(Z^+(t), Z^-(t))$, for each $t \geq cn^2$,

$$
E[d_H(Z^+(t+1), Z^-(t+1))] < \left(1 - \frac{\delta}{n^2}\right) E[d_H(Z^+(t), Z^-(t))] + e^{-\Omega(n)}.
$$

This allows us to control $d_{TV}(Z^+(t), Z^-(t))$ at $t = n^3$. Then, the sub-multiplicative property of $d_{TV}(\cdot, \cdot)$, implies that there exists a constant $\alpha > 0$ such that for each $t \geq n^3$,

$$
d_{TV}(Z^+(t), Z^-(t)) \leq e^{-\alpha t / n^2}.
$$

This and standard relations between total variation distance and the spectral gap then allows one to quickly deduce that the relaxation time is $O(n^2)$. The lower bound follows by proving an upper bound on the spectral gap using the variational formula (20) and a suitable test function.

3.2. Gaussian concentration. Theorem 1 is the main result of this paper and the key tool we rely on, is a version of Stein’s method for concentration via construction of suitable exchangeable pairs, developed in [8, 10] (see Section 5.1 for details). This combined with temporal mixing results derived from [5] allows us to obtain the sought bounds. Informally, to make this strategy work, one has to control the $L^\infty$-norm of the function

$$
F(x, y) := \sum_{t=0}^{\infty} (P_t f(x) - P_t f(y)).
$$

In particular, for our application, it will suffice to consider $y$ which differs from $x$ only at one edge, say $e$. Thus, to control the RHS in (31), using the Lipschitz-ness of $f$, with $x, y$ as above, it suffices to bound the $L_1$ norm of the vector

$$
r(t) = (P(Z^x(t) - Z^y(t))_{e \in E(K_n)}.
$$

for the monotone coupling $(Z^x(t), Z^y(t))_{t \geq 0}$. More precisely, we will show the following bounds for $\|r(t)\|_1$ which will allow us to bound the sum on the RHS of (31): for some $C, d, \alpha > 0$,

$$
\|r(t)\|_1 \leq \begin{cases}
C & t \leq dn^2, \\
(1 - \frac{\delta}{n^2})^t d n^2 & dn^2 \leq t \leq n^3, \\
e^{-\alpha t / n^2} & t \geq n^3.
\end{cases}
$$

The first one is a crude bound derived from straightforward considerations, the second bound follows from a contraction result deduced from the estimates in [5], while the third bound follows from the second by sub-multiplicative properties of total variation distance.

3.3. Central limit theorem. The strategy to prove the central limit theorem is to estimate the correlation structure of the edge variables $X_e$. This is a rather delicate task and in particular, we can obtain useful estimates only when the edges are vertex disjoint. More precisely, we prove the following key $k$-correlation estimate: if $k$ many distinct edges $i_1, \ldots, i_k$ do not share a vertex, then

$$
|\mathbb{E}[(X_{e_{i_1}} - \mathbb{E}X_{e_{i_1}}) \cdots (X_{e_{i_k}} - \mathbb{E}X_{e_{i_k}})]| = O\left(\frac{1}{n^{k/2}}\right).
$$

(33)
In order to establish the above bound, we first obtain quantitative estimates on the total variation distance between the distribution of an edge variable conditioned on other edges and the unconditional marginal (Proposition 6.2) as applications of the Gaussian concentration result and FKG inequality. In fact, we show that the fluctuation arising from conditioning on a fixed number of other edges is at most \( O\left(\frac{1}{n}\right) \):

\[
\left| \mathbb{P}(X_{e_1} = 1) - \mathbb{P}(X_{e_1} = 1|X_{e_2} = a_1, \ldots, X_{e_k} = a_k) \right| = O\left(\frac{1}{n}\right).
\] (34)

The \( k \)–correlation estimate is then proved using the above result and a conditional-version of Theorem 1. This allows us to complete the proof using a moment method argument.

3.4. **Bounding \( W_1 \)–Wasserstein distance:** The proof proceeds by constructing a coupling of the stationary GD on the ERGM and \( G(n,p^*) \) respectively, where the same edge is updated in the two Markov chains and the update probabilities are coupled in an optimal way. We then show as an application of the Gaussian concentration result, that with high probability the total variation distance between the update probabilities of the updated edge in the two models is no more than \( O\left(\frac{\log n}{n}\right) \), throughout the time it takes to update all the edges. The above then implies that the amount of discrepancy induced between the two chains till all the edges are updated is no more than \( O(\sqrt{\log n}) \), which along with stationarity of the chains is enough to finish the proof.

3.5. **Organization of the paper.** Several general facts about ERGM in the high temperature phase are reviewed in Section 4; while Dobrushin’s uniqueness regime is highlighted in Section 4.1, the general high temperature regime is discussed in Section 4.2. The short proof of the Poincaré inequality appears towards the end of this section.

The proof of main result in this article establishing Gaussian concentration for Lipschitz functions appears in Section 5. The subsequent Sections 6 and 7 contain the proofs of the central limit theorem (Theorem 2) and the bound on the Wasserstein distance to Erdős-Rényi graphs (Theorem 3) respectively.

3.6. **Acknowledgement.** The authors thank Charles Radin and Arthur Sinulis for useful comments. They also thank two anonymous referees for their detailed and thoughtful comments that helped improve the paper. SG’s research is partially supported by a Sloan Research Fellowship in Mathematics and NSF Award DMS-1855688. KN’s research is supported by a summer grant of the UC Berkeley Mathematics department.

4. **ERGM in the high temperature regime**

To motivate some of our arguments and set up further necessary notation, we review some useful facts about the ERGM in the high temperature regime. We begin by considering the perturbative case [18, 19].

4.1. **Dobrushin’s uniqueness regime.** For edges \( e \) and \( f \), let

\[
a_{e,f} := \sup_{x \in \mathcal{G}_n} d_{TV}(\pi_e(\cdot|x_{f+}),\pi_e(\cdot|x_{f-})).
\] (35)

The **Dobrushin interdependent matrix** \( A \), of size \( M \times M \), is defined as \( A = (a_{e,f})_{e,f \in E(K_n)} \). We say that the Gibbs measure \( \pi \) satisfies the \( L^2 \)-version of **Dobrushin’s uniqueness condition** if the matrix
A satisfies \( \|A\|_2 < 1 \). This condition is slightly different from the original \( L^1 \)-version of Dobrushin’s uniqueness condition, where the matrix \( A \) is assumed to satisfy \( \|A\|_1 < 1 \) (\( \|A\|_2 \) and \( \|A\|_1 \) denote the norms of \( A \), thought of as an operator from \( \mathbb{R}^M \) to itself equipped with the \( \ell_2 \) and \( \ell_1 \) norms respectively). Recalling (10) and the fact that the total variation distance between two probability measures \( \mu \) and \( \nu \) on the two spin system \( \{0,1\} \) is \( |\mu(\{1\}) - \nu(\{1\})| \), for any configuration \( x, \)

\[
d_{TV}(\pi_e(\cdot|x_f^+),\pi_e(\cdot|x_f^-)) = |\pi_e(x_e = 1|x_f^+) - \pi_e(x_e = 1|x_f^-)|
\]

\[
= \frac{1}{2} \left( 1 + \tanh(\partial_e H(x_f^+)/2) \right) - \frac{1}{2} \left( 1 + \tanh(\partial_e H(x_f^+)/2) \right)
\]

\[
\leq \frac{1}{4} |\partial_e H(x)| = \frac{1}{4} \sum_{i=1}^{s} \beta_i \frac{N_{G_i}(x,e,f)}{n |V_i| - 2},
\]

(36)

where the above notations were introduced around (4). Here, the last inequality above follows from the mean value theorem, the definition of \( \partial_e H(x) \) and that \( \sup_{x \geq 0} \frac{d}{dx} \tanh(x) = 1 \). The last equality in (36) follows from the following expressions of the discrete derivatives. Recalling the definitions from (2), we have

\[
\partial_e H(x) = \sum_{i=1}^{s} \beta_i \frac{N_{G_i}(x,e)}{n |V_i| - 2},
\]

(37)

and hence,

\[
\partial_{ef} H(x) = \partial_f(\partial_e H(x)) = \sum_{i=1}^{s} \beta_i \frac{\partial_f N_{G_i}(x,e)}{n |V_i| - 2} = \sum_{i=1}^{s} \beta_i \frac{N_{G_i}(x,e,f)}{n |V_i| - 2}.
\]

We now define an \( M \times M \) symmetric matrix \( L = (L_{ef}) \) by

\[
L_{ef} := \frac{1}{4} \sum_{i=1}^{s} \beta_i \frac{N_{G_i}(K_n,e,f)}{n |V_i| - 2}.
\]

(38)

Then, by (36), we have that for any edges \( e \) and \( f, \)

\[
a_{ef} \leq L_{ef}.
\]

(39)

Using the fact that

\[
\sum_{e:e\neq f} N_G(K_n,e,f) = (|E| - 1)N_G(K_n,f) = 2|E|(|E| - 1) \left( \frac{n - 2}{|V| - 2} \right) (|V| - 2)!,
\]

(40)

we conclude that

\[
\sum_{e:e\neq f} L_{ef} = \frac{1}{4} \sum_{i=1}^{s} \beta_i \frac{\sum_{e:e\neq f} N_{G_i}(K_n,e,f)}{n |V_i| - 2} < \frac{1}{2} \sum_{i=1}^{s} \beta_i |E_i|(|E_i| - 1).
\]

Since \( L \) is symmetric, this implies that

\[
\sup_{n} \|L\|_2 \leq \sup_{n} \|L\|_1 < \infty.
\]

(41)

Assume that \( \beta = (\beta_1, \ldots, \beta_s) \), satisfies the condition

\[
\frac{1}{2} \sum_{i=1}^{s} \beta_i |E_i|(|E_i| - 1) < 1.
\]

(42)
Then, by (39),
\[ \|A\|_2 \leq \|L\|_2 \leq \|L\|_1 < 1, \quad \|A\|_1 \leq \|L\|_1 < 1. \]

Thus, all \( \beta \) satisfying (42) lies in the \( (L^2\text{-version of}) \) DU regime. Although in such regimes, for classical spin systems on lattices, it is well known the stationary measure behaves qualitatively like a product measure in a rather strong sense, satisfying the LSI and other related concentration of measure properties (see e.g. [7, 31, 50, 49, 52, 38] for more details), for the ERGM, as mentioned earlier such an analysis was carried out only very recently in [46].

On the other hand, the \( L^1\)-version of DU condition, \( \|A\|_1 < 1 \), implies that there exists a coupling for the GD such that the Hamming distance strictly contracts. Although this is a well-known fact, we include the brief proof for the sake of completeness. Suppose that two configurations \( x, y \in G_n \) differ at edges \( j_1, \cdots, j_l \), and take a path of configurations \( x = w^0, w^1, \cdots, w^{l-1}, w^l = y \) such that \( w^{j-1} \) and \( w^j \) differ only at the edge \( j_i \). Then, under the grand coupling \( (Z^x(t), Z^y(t)) \),
\[
\mathbb{P}(Z^x_e(1) \neq Z^y_e(1)) = \left( 1 - \frac{1}{M} \right) \mathbb{1}\{x_e \neq y_e\} + \frac{1}{M} d_{TV}(\pi_e(\cdot|x), \pi_e(\cdot|y)) \leq \left( 1 - \frac{1}{M} \right) \mathbb{1}\{x_e \neq y_e\} + \frac{1}{M} \sum_{i=1}^l d_{TV}(\pi_e(\cdot|w^{i-1}), \pi_e(\cdot|w^i)) \leq \left( 1 - \frac{1}{M} \right) \mathbb{1}\{x_e \neq y_e\} + \frac{1}{M} \sum_{i=1}^l a_{ej_i} \mathbb{1}\{x_e \neq y_e\} + \frac{1}{M} \sum_{f:f \neq e} a_{ef} \mathbb{1}\{x_f \neq y_f\}. \tag{43}
\]

Adding these inequalities over all edges \( e \), we obtain
\[
\mathbb{E}[d_H(Z^x(1), Z^y(1))] \leq \left( 1 - \frac{1}{M} \right) d_H(x, y), \tag{44}
\]
proving a strict contraction in the Hamming distance if \( \|A\|_1 < 1 \). This also implies bounds on the spectral gap (see the monograph [34] for details).

Note that in general, by (39) and (44), for any two configurations \( x, y \in G_n \), under the grand coupling \( (Z^x(t), Z^y(t)) \),
\[
\mathbb{E}[d_H(Z^x(1), Z^y(1))] \leq \left( 1 - \frac{1}{M} \right) d_H(x, y),
\]

which by induction implies
\[
\mathbb{E}[d_H(Z^x(t), Z^y(t))] \leq \left( 1 - \frac{1}{M} \right)^t d_H(x, y). \tag{45}
\]

**Remark 4.1.** It is proved in [14, Theorem 6.2] that when \( \beta \) satisfies (42), the model is in the replica symmetric phase even for negative values of \( \beta \), i.e., the maximizers in (15) are constant functions provided that \( \frac{1}{2} \sum_{i=1}^s |\beta_i| |E_i| (|E_i| - 1) < 1 \). In fact, under the same condition which falls in the DU regime, as mentioned earlier, [46] established the LSI and derived concentration of measure properties.
4.2. Sub-critical but beyond Dobrushin’s uniqueness regime. Even beyond DU regime, where strict contraction property of GD no longer holds, it was nonetheless proved in [5] that there exists a set $T \subseteq \mathcal{G}_n$, such that

(1) GD starting from two states in $T$ exhibits a strict contraction.

(2) From any starting state, GD hits $T$ with high probability within $O(n^2)$ steps.

We will need precise versions of the above statements to prove Theorem 2.1. To this end, it would be convenient to introduce the notion of normalized subgraph counting number, following [5]. For any configuration $x$, edge $e$, and a graph $G$, define $r_G(x,e)$ by

$$r_G(x,e) := \left( \frac{N_G(x,e)}{2|E|n^{|V|-2}} \right)^{1/|\{y\}|},$$

and then define

$$r_{1,\text{max}}(x) := \max_{e,G \in \mathcal{G}_a} r_G(x,e), \quad r_{1,\text{min}}(x) := \min_{e,G \in \mathcal{G}_a} r_G(x,e),$$

where $\mathcal{G}_a$ denotes a collection of all graphs with at most $a$ many vertices. Recall from (8) that $a$ is the number satisfying $|V_i| \leq a$ for $i = 1, 2, \cdots, s$. Recalling $p^*$ from (20), and $\delta > 0$, let $\mathcal{T}_{\delta}$ be the collection of configurations defined by

$$\mathcal{T}_{\delta} := \{x \in \mathcal{G}_n : p^* - \delta < r_{1,\text{min}}(x) \leq r_{1,\text{max}}(x) < p^* + \delta\}.$$ 

The following is one of the key results in [5].

**Theorem 4.2.** [5, Lemma 16] Suppose that $\beta$ lies in the high temperature regime. Then, for any $\delta > 0$, there exists a constant $c$ such that for any initial configuration $x \in \mathcal{G}_n$, when $t \geq cn^2$,

$$\mathbb{P}(Z^e(t) \in \mathcal{T}_{\delta}) \geq 1 - e^{-\Omega(n)}.$$ 

Now, we introduce a generalized version of the Dobrushin’s matrix defined in Section 4.1, where instead of taking a supremum over all configurations in (35), we keep the dependence on the configuration $x$ to define $A(x) := (a_{ef}(x))_{e,f \in E(K_n)}$ by,

$$a_{ef}(x) := d_{TV}(\pi_e(\cdot|xf_+), \pi_e(\cdot|xf_-)).$$

Then, by definition,

$$a_{ef}(x) = \Phi(\partial_e H(xf_+)) - \Phi(\partial_e H(xf_-)).$$

For two configurations $y \leq z$, consider the monotone coupling of $Z^y(t)$ and $Z^z(t)$, and define the $M$-dimensional vector $r(t) = (r(t)_e)_{e \in E(K_n)}$ by

$$r(t)_e = \mathbb{P}(Z^y_e(t) \neq Z^z_e(t)).$$

The proof of [5, Lemma 18] implies the following proposition, which states the pointwise decay of the vector $r(t)$ along the GD starting from the set $\mathcal{T}_{\delta}$.

**Proposition 4.3.** Suppose that $\beta$ lies in the high temperature phase. Then, there exists a constant $\delta > 0$ such that the following holds for sufficiently large $n$. For any $y \leq z$ with $y, z \in \mathcal{T}_{\delta}$, there is a $M \times M$ matrix $K = K(n,y,z)$ with

$$||K||_1 < 1 - \frac{\delta}{M}.$$
and
\[ r(1) \leq Kr(0), \] (51)
where the inequality is pointwise.

Since neither the statement or its proof explicitly appears in [5], we include the details for completeness.

Proof. The proof follows from the following result: if \( \beta \) lies in the high temperature phase, then there exists sufficiently small \( \delta > 0 \) such that for sufficiently large \( n \):
\[ x \in T_\delta \Rightarrow \|A(x)\|_1 < 1 - \delta. \] (52)
To see the above, by mean value theorem, for some \( \partial_e H(x_f-) \leq u_0 \leq \partial_e H(x_f+) \),
\[ a_{ef}(x) = \Phi(\partial_e H(x_f+)) - \Phi(\partial_e H(x_f-)) = \Phi'(u_0)(\partial_e H(x_f+) - \partial_e H(x_f-)). \] (53)
Note that for any edges \( e \) and \( f \),
\[ r_G(x_f+, e) - r_G(x, e) = o(1), \quad r_G(x, e) - r_G(x_f-, e) = o(1). \] (54)
In fact, the quantity \( N_G(x_f+, e) - N_G(x, e) \) is \( O(n|V(G)|^{-3}) \) if \( e \) and \( f \) share an edge and \( O(n|V(G)|^{-4}) \) if \( e \) and \( f \) do not share an edge (similar property holds for \( N_G(x, e) - N_G(x_f-, e) \) as well). Using the inequality \( (z_1 + z_2)^p - z_1^p \leq z_2^p \) for \( z_1, z_2 \geq 0, \ 0 < p \leq 1 \), and recalling the normalization \( n|V(G)|^{-2} \) in (46), we have (54). Hence, since \( x \in T_\delta \), for sufficiently large \( n \), we have \( x_f+, x_f- \in T_{2\delta} \) for any edge \( f \).

Thus recalling (17), and using
\[ N_G(K_n, g) = 2|E|[\frac{n-2}{|V|-2}](|V|-2)! = (1 - o(1))2|E|n|V_i|-2 \leq 2|E|n|V_i|-2, \] (55)
we obtain,
\[ \partial_e H(x_f+) \leq \sum_{i=1}^{s} \frac{\beta_i(p^* + 2\delta)|E_i|^{-1}N_G_i(K_n, e)}{n|V_i|-2} \leq \Psi'_\beta(p^* + 2\delta), \] (55)
and
\[ \partial_e H(x_f+) \geq \sum_{i=1}^{s} \frac{\beta_i(p^* - 2\delta)|E_i|^{-1}N_G_i(K_n, e)}{n|V_i|-2} = (1 - o(1))\Psi'_\beta(p^* - 2\delta). \]
Since the same inequalities also hold for \( \partial_e H(x_f-) \), using (53) and the fact that \( \Phi' \) is decreasing on \( (0, \infty) \), for any \( \eta > 0 \), for sufficiently large \( n \) and small \( \delta > 0 \),
\[ a_{ef}(x) \leq (1 + \eta)\Phi'(\Psi'_\beta(p^*))(\partial_e H(x_f+) - \partial_e H(x_f-)) \]
\[ = (1 + \eta)\Phi'(\Psi'_\beta(p^*)) \sum_{i=1}^{s} \frac{N_G_i(x, e, f)}{n|V_i|-2}. \] (56)
Since \( \varphi'_\beta(p^*) < 1 \) (see (19)), one can take constants \( \eta, \delta > 0 \) sufficiently small satisfying
\[ (1 + \eta)^2 \varphi'_\beta(p^*) < 1 - \delta. \] (57)
For such $\eta$, define a $M \times M$ symmetric matrix $U(x) = (U_{ef}(x))$ by
\[
U_{ef}(x) := (1 + \eta)\Phi'(\Psi_B(p^*)) \sum_{i=1}^s \beta_i \frac{N_{G_i}(x, e, f)}{n|V_i| - 2}.
\] (58)

Then, by (56), for sufficiently large $n$ and sufficiently small $\delta > 0$,
\[
a_{ef}(x) \leq U_{ef}(x).
\] (59)

We verify that the condition (57) ensures that
\[
\|U(x)\|_1 < 1 - \delta.
\] (60)

In fact, denoting $G_g$ by a graph obtained from $G$ by removing an edge $g$, since $x \in T_\delta$, using [5, Lemma 10] and (40), for sufficiently small $\delta > 0$,
\[
\sum_{e: e \neq f} \sum_{i=1}^s \frac{\beta_i N_{G_i}(x, e, f)}{n|V_i| - 2} = \sum_{i=1}^s \beta_i \frac{\sum_{g \in E(G_i), g \neq f} N_{G_i}(x, f)}{n|V_i| - 2}
\leq \sum_{i=1}^s \beta_i \frac{\sum_{g \in E(G_i), g \neq f} (p^* + \delta)^|E_i| - 2 N_{G_i}(K_n, f)}{n|V_i| - 2}
= \sum_{e: e \neq f} \sum_{i=1}^s \beta_i (p^* + \delta)^|E_i| - 2 \frac{N_{G_i}(K_n, e, f)}{n|V_i| - 2}
\leq (1 + \eta) \sum_{i=1}^s \beta_i (p^*)^{|E_i| - 2} |E_i| (|E_i| - 1).
\]

Applying this to (58), by (57), we have
\[
\sum_{e: e \neq f} U_{ef}(x) \leq (1 + \eta)^2 \Phi'(\Psi_B(p^*)) \sum_{i=1}^s 2\beta_i |E_i| (|E_i| - 1)(p^*)^{|E_i| - 2}
= (1 + \eta)^2 \Phi'(\Psi_B(p^*)) \Psi_B(p^*) = (1 + \eta)^2 \Phi'(p^*) < 1 - \delta.
\] (61)

This concludes the proof of (60). Therefore, by (59) and (60), we have (52).

Equipped with (52), one can conclude the proof of Proposition 4.3. Suppose that two configurations $y \leq z$ with $d_H(y, z) = \ell$, differ at coordinates $i_1, \cdots, i_\ell$. Then, consider a sequence
\[
y = w^0 \leq w^1 \leq \cdots \leq w^{\ell-1} \leq w^\ell = z,
\]
where $w^k = w^{k-1} \cup i_k$. Since $y, z \in T_\delta$ and $r_G(x, e)$ is an increasing function in $x$, we have $w^k \in T_\delta$ for all $k = 1, \cdots, \ell$. Note that using a triangle inequality and the definition of Dobrushin Matrix (49),
\[
d_{TV}(\pi_e(\cdot | y), \pi_e(\cdot | z)) \leq \sum_{k=1}^\ell d_{TV}(\pi_e(\cdot | w^{k-1}), \pi_e(\cdot | w^k)) \leq \sum_{k=1}^\ell a_{eik}(w^k).
\]

Hence,
\[
\mathbb{P}(Z_e^y(1) \neq Z_e^z(1)) = \left(1 - \frac{1}{M}\right) \mathbb{1}\{y_e \neq z_e\} + \frac{1}{M} d_{TV}(\pi_e(\cdot | y), \pi_e(\cdot | z))
\]
\[ \leq \left( 1 - \frac{1}{M} \right) \mathbb{1}_{\{ y_e \neq z_e \}} + \frac{1}{M} \sum_{k=1}^{l} a_{eik}(w^k). \] (62)

Define a matrix \( K' = (K'_{ef})_{e,f \in E(K_n)} \) in the following way: for any \( e \) and \( 1 \leq k \leq l \), let \( K'_{eik} := a_{eik}(w^k) \), and otherwise let \( K'_{ef} = 0 \) for \( f \neq i_1, \ldots, i_l \). Since \( w^k \in \mathcal{T}_3 \), by (52), we have \( \|K'\|_1 < 1 - \delta \) for sufficiently small \( \delta > 0 \). This is because \( \|K'\|_1 \) is nothing but the maximum \( L_1 \) norm of its columns. Now, by definition, columns of \( K' \) indexed by edges other than \( i_1, i_2, \ldots, i_k \) are zero. Further, for any \( 1 \leq \ell \leq k \), the column of \( K' \) indexed by \( i_\ell \) agrees with the corresponding column of \( A(w^\ell) \), which by (52), has \( L_1 \) norm bounded by \( 1 - \delta \).

Thus, the matrix

\[ K = \left( 1 - \frac{1}{M} \right) I + \frac{1}{M} K' \]

satisfies \( \|K\|_1 < 1 - \frac{\delta}{M} \). Now writing (62) for every edge \( e \) in matrix form yields (51).

\[ \square \]

We next state a straightforward consequence of Proposition 4.3 which will be crucial in the proof of the Gaussian concentration result, which states that GD starting from configurations in the set \( T_\delta \) with small \( \delta > 0 \) exhibits contraction. Formally, if \( \beta \) lies in the high temperature phase, then there exists sufficiently small \( \delta > 0 \) such that for \( y \leq z \) with \( y, z \in \mathcal{T}_3 \), under the grand coupling,

\[ \mathbb{E}[d_H(Z^y(1), Z^z(1))] \leq \left( 1 - \frac{\delta}{M} \right) d_H(y, z). \] (63)

Therefore, by (63) and Theorem 4.2, for any configurations \( y \leq z \), under the grand coupling, for \( t \geq cn^2 \),

\[ \mathbb{E}[d_H(Z^y(t+1), Z^z(t+1))] \leq \left( 1 - \frac{\delta}{M} \right) \mathbb{E}[d_H(Z^y(t), Z^z(t))] + e^{-\Omega(n)}. \]

It follows that for \( t \geq cn^2 \),

\[ \mathbb{E}[d_H(Z^y(t), Z^z(t))] \leq \left( 1 - \frac{\delta}{M} \right)^{t-cn^2} \mathbb{E}[d_H(Z^y(cn^2), Z^z(cn^2))] + e^{-\Omega(n)}. \] (64)

In particular, recalling that \( Z^+(t) \) and \( Z^-(t) \) denote the GD starting from the complete and empty initial configurations respectively, under the grand coupling, for \( t \geq cn^2 \),

\[ \mathbb{P}(Z^+(t) \neq Z^-(t)) \leq \mathbb{E}[d_H(Z^+(t), Z^-(t))] \leq \left( 1 - \frac{\delta}{M} \right)^{t-cn^2} M + e^{-\Omega(n)}. \] (65)

Since ERGM with positive \( \beta \) is a monotone system, by above, we have

\[ \sup_{\mu, \nu \in \mathcal{P}(G_n)} d_{TV}(\mu_t, \nu_t) \leq \left( 1 - \frac{\delta}{M} \right)^{t-cn^2} M + e^{-\Omega(n)}. \] (66)

The above preparation already allows us to finish the short proof of the Poincaré inequality.

**Proof of Theorem 2.1.** The upper bound follows from the variation characterization (20) and plugging in the test function \( f : G_n \rightarrow \mathbb{R} \) where \( f(x) = x_e - \mathbb{E}(X_e) \), where \( e \) is a fixed edge and \( X \) is distributed according to ERGM(\( \beta \)). To see this, note that for such a function, we have \( (f(x^e) - f(x))^2 = 1 \) and \( (f(x^{e'}) - f(x))^2 = 0 \) for any \( e' \neq e \), where recall from (11) that \( x^e \) denotes
the configuration obtained from $x$ by flipping the state of $e$. Thus, using $c(x, e) \leq 1$, we have
\[
\mathcal{E}(f, f) = \frac{1}{M} \sum_{x \in \mathcal{G}_n} c(x, e)\pi(x) \leq \frac{1}{M} \leq \frac{C}{n^2},
\] 
(67)
In addition, by Remark 6.3 stated later, which gives the asymptotics $\lim_{n \to \infty} P(X_e = 1) = p^*$ in the high temperature regime, one can deduce that $\int f^2 d\pi = \mathbb{E}(X_e^2) - (\mathbb{E}X_e)^2 \xrightarrow{n \to \infty} p^*(1 - p^*)$ and hence is uniformly bounded away from 0. This combined with (67) gives a desired upper bound of order $\frac{1}{n^2}$ for the spectral gap.

For the lower bound, by the following well known relation between total variation distance and spectral gap (see [34, equation (12.15)] for more details),
\[
(1 - \gamma)^t \leq 2d_{TV}(Z_t(x), \pi),
\]
it suffices to prove that there exists $d > 0$ such that for sufficiently large $n$ and any $t \geq n^3$, there is a coupling of $Z^+(t)$ and $Z^-(t)$ satisfying
\[
P(Z^+(t) \neq Z^-(t)) \leq e^{-dt/n^2}.
\] 
(68)
The proof is now complete by noticing that the above is a straightforward consequence of sub-multiplicative property of total variation distance and plugging in $t = n^3$ in (66). □

In the next section, using the contraction estimates we prove the Gaussian concentration.

5. Gaussian concentration: Proof of Theorem 1

As indicated in Section 3, Stein’s method to prove concentration will be the key tool we rely on. We start by briefly reviewing the pertinent theory of the latter.

5.1. Stein’s method for concentration. In [10], Chatterjee introduced a beautiful new argument showing how Stein’s method of exchangeable pairs can be used to obtain concentration results and presented several applications in [8, 13]. For instance, a concentration result for the magnetization (say $m$) of the Curie-Weiss model with an inverse temperature $\beta$ and external field $h$ was obtained by showing, for any $\beta \geq 0$, with high probability, $m \approx \tanh(\beta m + \beta h)$. Recall that a pair $(X, X')$ is said to be exchangeable if $(X, X')$ and $(X', X)$ have the same distribution. The following is the key theorem we will use as input.

**Theorem 5.1.** [10, Theorem 1.5] For a separable metric space $S$, let $(X, X')$ be an exchangeable pair of random variables taking values in $S$. Suppose that $f : S \to \mathbb{R}$ and an antisymmetric function $F(x, y) : S \times S \to \mathbb{R}$ are square-integrable, and satisfies
\[
f(X) := \mathbb{E}[F(X, X')|X].
\]
Then, define a function $g : S \to \mathbb{R}$ by
\[
g(x) := \frac{1}{2}\mathbb{E}[|(f(X) - f(X'))F(X, X')||X = x].
\]
If $|g(x)| \leq C$, then for any $\theta \in \mathbb{R}$,
\[
\mathbb{E}e^{\theta f(X)} \leq e^{C\theta^2/2}.
\] 
(71)
In particular, for any \( t \geq 0 \),
\[
\mathbb{P}(\|f(X)\| \geq t) \leq 2e^{-t^2/2C}. \tag{72}
\]

We now describe the construction of the anti-symmetric function to be used. For an exchangeable pair \((X, X')\), there is a natural way to associate a reversible Markov kernel \( \mathcal{P} \):
\[
\mathcal{P}f(x) = \mathbb{E}[f(X')|X = x]. \tag{73}
\]
When we are given a function \( f: S \to \mathbb{R} \) such that \( \mathbb{E}f(X) = 0 \) and
\[
\sup_{x, y \in S} \sum_{k=0}^{\infty} |\mathcal{P}^k f(x) - \mathcal{P}^k f(y)| < \infty,
\]
then it is easy to check that the function
\[
F(x, y) := \sum_{k=0}^{\infty} (\mathcal{P}^k f(x) - \mathcal{P}^k f(y)) \tag{74}
\]
is antisymmetric and satisfies the relation (69) (see [8, Chapter 4] for details). We are now ready to prove Theorem 1.

**Proof of Theorem 1.** Without loss of generality, let us assume that \( \mathbb{E}_n f = 0 \). Let \( X = Z(0) \sim \pi \) and \( X' = Z(1) \) be the step 1 distribution of GD starting from \( X \). By reversibility of \( P \) (see (73)), \( (X, X') \) is an exchangeable pair. Let us define the antisymmetric function \( F \) via (74), and subsequently the function \( g \) as in (70). The first order of business is to obtain an \( L^\infty \) bound on the function \( g \). We will in fact prove that
\[
2|g(x)| \leq C_0 \|\nu\|_1 \|\nu\|_\infty, \tag{75}
\]
for some constant \( C_0 > 0 \) independent of \( n \). This in conjunction with Theorem 5.1, completes the proof of the theorem. The rest of the proof will be devoted to verifying (75). By (70)
\[
2|g(x)| = \mathbb{E}\left[ ((f(X) - f(X'))F(X, X')|X = x \right]
\leq \frac{1}{M} \sum_{l=1}^{M} |(f(x_{l+}) - f(x_{l-}))F(x_{l+}, x_{l-})|. \tag{76}
\]
For each edge \( l \), let us obtain an upper bound of \( |(f(x_{l+}) - f(x_{l-}))F(x_{l+}, x_{l-})| \). For any configuration \( x \) and edge \( l \), let \( X_{l+}(t) = Z_{x_{l+}}(t) \) and \( X_{l-}(t) = Z_{x_{l-}}(t) \) be the GD starting from the initial configurations \( x_{l+} \) and \( x_{l-} \), respectively. Then, since \( f \) is \( \nu \)-Lipschitz, we have
\[
|f(x_{l+}) - f(x_{l-})||F(x_{l+}, x_{l-})| \leq v_l \sum_{t=0}^{\infty} |\mathbb{E}f(X_{l+}^{t+}(t)) - \mathbb{E}f(X_{l-}^{t-}(t))|.
\]
Note that for each \( t \) and any coupling of \( X_{l+}^{t+}(t) \) and \( X_{l-}^{t-}(t) \),
\[
|\mathbb{E}f(X_{l+}^{t+}(t)) - \mathbb{E}f(X_{l-}^{t-}(t))| \leq \sum_{j=1}^{M} v_j \mathbb{P}(X_{j+}^{l+}(t) \neq X_{j-}^{l-}(t)),
\]

By the sub-critical nature of the exponential random graphs, this implies that
\[
|\mathbb{E}f(X_{l+}^{t+}(t)) - \mathbb{E}f(X_{l-}^{t-}(t))| \leq 2e^{-t^2/2C}.
\]

By taking the limit as \( t \to \infty \), we obtain
\[
|g(x)| \leq \frac{1}{M} \sum_{l=1}^{M} 2e^{-t^2/2C} \leq C_0 \|\nu\|_1 \|\nu\|_\infty,
\]

which completes the proof of Theorem 1.
which implies that
\[
|f(x_{l+}) - f(x_{l-})| \leq v_l \sum_{t=0}^{\infty} \sum_{j=1}^{M} v_j \mathbb{P}(X_j^{l,+}(t) \neq X_j^{l,-}(t)).
\] (77)

We will couple \(X^{l,+}(t)\) and \(X^{l,-}(t)\) through the monotone coupling. For such a coupling, let \(r(x, l, t)\) be the \(M\)-dimensional vector with \(r(x, l, t)_i := \mathbb{P}(X_j^{l,+}(t) \neq X_j^{l,-}(t))\). Then,
\[
\sum_{j=1}^{M} v_j \mathbb{P}(X_j^{l,+}(t) \neq X_j^{l,-}(t)) \leq \|r(x, l, t)\|_1 \|v\|_\infty.
\] (78)

Since \(\|L\|_1\) (recall that \(L\) is defined in (38)) is bounded above by a constant independent of \(n\) (see (41)), using (45) and the fact \(\|r(x, l, 0)\|_1 = 1\), there exists a constant \(C > 0\) such that for \(t \leq cn^2\),
\[
\|r(x, l, t)\|_1 \leq \left(1 - \frac{1}{M} + \frac{\|L\|_1}{M}\right) t \|r(x, l, 0)\|_1 \leq C.
\]

Thus, it follows by (78) that
\[
t \leq cn^2 \Rightarrow \sum_{j=1}^{M} v_j \mathbb{P}(X_j^{l,+}(t) \neq X_j^{l,-}(t)) \leq C \|v\|_\infty.
\] (79)

Also, since \(\|r(x, l, cn^2)\|_1 \leq C\), by (64) and the fact
\[
\mathbb{E}[d_H(X^{l,+}(t), X^{l,-}(t))] = \mathbb{E}\left[\sum_{j=1}^{M} \mathbb{1}\{X_j^{l,+}(t) \neq X_j^{l,-}(t)\}\right] = \|r(x, l, t)\|_1,
\]
for \(cn^2 < t < n^3\),
\[
\sum_{j=1}^{M} v_j \mathbb{P}(X_j^{l,+}(t) \neq X_j^{l,-}(t)) \leq \|r(x, l, t)\|_1 \|v\|_\infty \leq \left(1 - \frac{\delta}{M}\right) t^{-cn^2} C + e^{-\Omega(n)} \|v\|_\infty.
\] (80)

Now note that for \(t \geq n^3\),
\[
\mathbb{P}(X^{l,+}(t) \neq X^{l,-}(t)) \leq e^{-\alpha t/n^2}
\] (81)
for some \(\alpha > 0\). The above bound follows by plugging in \(t = n^3\) in (65) and using the sub-multiplicative property of the quantity \(\sup_{\mu, \nu \in \mathcal{P}(G_n)} d_{TV}(\mu_t, \nu_t)\) along with monotonicity (see (66)).

By (78) and (81),
\[
t \geq n^3 \Rightarrow \sum_{j=1}^{M} v_j \mathbb{P}(X_j^{l,+}(t) \neq X_j^{l,-}(t)) \leq n^2 e^{-\alpha t/n^2} \|v\|_\infty \leq e^{-\alpha' t/n^2} \|v\|_\infty
\] (82)

for some \(\alpha' > 0\) chosen slightly smaller than \(\alpha\) to absorb the \(n^2\) pre-factor.

Therefore, applying (79), (80) and (82) to (77),
\[
|f(x_{l+}) - f(x_{l-})| \leq v_l \sum_{t=0}^{\infty} \sum_{j=1}^{M} v_j \mathbb{P}(X_j^{l,+}(t) \neq X_j^{l,-}(t))
\]
\[ \leq v_l \| v \|_\infty \left( \sum_{t=0}^{cn^2} C + \sum_{t=cn^2+1}^{n^3-1} \left( (1 - \frac{\delta}{M}) t^{-cn^2} C + e^{-\Omega(n)} \right) + \sum_{t=n^3}^{\infty} e^{-\alpha't/n^2} \right) \]
\[ \leq v_l \| v \|_\infty O(n^2). \]

Recalling that \( M \) is of order \( n^2 \), this implies that for some constant \( C_0 > 0 \),
\[ 2|g(x)| \leq \frac{1}{M} \sum_{i=1}^{M} |f(x_{l+}) - f(x_{l-})||F(x_{l+}, x_{l-})| \leq C_0 \| v \|_1 \| v \|_\infty, \]
which in conjunction with Theorem 5.1, implies that for \( t \geq 0 \),
\[ P(|f(X)| \geq t) \leq 2e^{-t^2/C_0 \| v \|_1 \| v \|_\infty}. \]  
(83)

\[ \square \]

6. CENTRAL LIMIT THEOREM: PROOF OF THEOREM 2

The proof is at a very high level based on the moment method and will rely on delicate and novel estimates of the \( k \)-correlation functions of the form, \( E[X_{i_1} \cdots X_{i_k}] - (EX_{i_1}) \cdots (EX_{i_k}) \) for edges \( i_1, i_2, \ldots, i_k \). We first start with a few related results that will be used.

**Lemma 6.1.** For any fixed positive integer \( k \) and distinct edges \( i_1, \ldots, i_k \),
\[ |E[X_{i_1} \cdots X_{i_k}] - (EX_{i_1}) \cdots (EX_{i_k})| = O\left( \frac{1}{n} \right). \]  
(84)

**Proof.** The case \( k = 2 \) easily follows from the Gaussian concentration. In fact, since the Lipschitz vector of the function \( X_1 + \cdots + X_M \) is \((1, \ldots, 1)\), Theorem 1 implies that
\[ \text{Var}(X_1 + \cdots + X_M) = O(M). \]  
(85)

By symmetry, one can deduce that if two edges \( i \) and \( j \) does not share a vertex, then
\[ \text{Cov}(X_i, X_j) = O\left( \frac{1}{n^2} \right), \]  
(86)

whereas if two distinct edges \( i \) and \( j \) share a vertex, then
\[ \text{Cov}(X_i, X_j) = O\left( \frac{1}{n} \right). \]  
(87)

Above we use the fact that Cov\((X_i, X_j) \geq 0 \) by FKG inequality. In the case \( k \geq 3 \), we will crucially rely on the following estimate.

**Fact** [40, Equation (12)]: If random variables \( (Z_1, \ldots, Z_m) \) satisfy the positive quadrant dependent condition, i.e., for all \( x, y \in \mathbb{R} \),
\[ P(Z_i > x, Z_j > y) \geq P(Z_i > x)P(Z_j > y), \]  
(88)

then there exists a constant \( C > 0 \) such that for any \( C^1 \) functions \( f, g \) with bounded partial derivatives,
\[ |\text{Cov}(f(Z_1, \ldots, Z_m), g(Z_1, \ldots, Z_m))| \leq C \sum_{i=1}^{m} \sum_{j=1}^{m} \left\| \frac{\partial f}{\partial z_i} \right\|_\infty \left\| \frac{\partial g}{\partial z_j} \right\|_\infty \text{Cov}(Z_i, Z_j). \]  
(89)
Since \((X_{i_1}, \cdots, X_{i_j})\) satisfy \((88)\) by FKG inequality, using \((89)\) with \((86)\) and \((87)\),

\[
|\text{Cov}(X_{i_1} \cdots X_{i_{j-1}}, X_{i_j})| = O\left(\frac{1}{n}\right)
\]

for any \(j\) (this can be done by choosing \(g(x) = x\) and \(f(x_1, \cdots, x_{j-1}) = \varphi(x_1, \cdots, x_{j-1})x_1 \cdots x_{j-1}\) with a compactly supported smooth function \(\varphi\) satisfying \(\varphi = 1\) on \([-2, 2]^{j-1}\)). Using this fact repeatedly for \(1 \leq j \leq k\), we obtain \((84)\). \(\square\)

Next, we derive the following quantitative independence result, which will be a key ingredient to obtain \(k\)-correlation estimates.

**Proposition 6.2.** For any fixed positive integer \(k\) and \(a_1, \cdots, a_k, b_1, \cdots, b_k \in \{0, 1\}\), the following statement holds. If the edge \(j\) does not share a vertex with edges \(i_1, \cdots, i_k\), then

\[
\left| \mathbb{P}(X_j = 1|X_{i_1} = a_1, \cdots, X_{i_k} = a_k) - \mathbb{P}(X_j = 1|X_{i_1} = b_1, \cdots, X_{i_k} = b_k) \right| = O\left(\frac{1}{n^2}\right).
\]

Whereas, if the edge \(j\) shares vertices with some of the edges \(i_1, \cdots, i_k\), then

\[
\left| \mathbb{P}(X_j = 1|X_{i_1} = a_1, \cdots, X_{i_k} = a_k) - \mathbb{P}(X_j = 1|X_{i_1} = b_1, \cdots, X_{i_k} = b_k) \right| = O\left(\frac{1}{n}\right).
\]

**Remark 6.3.** A qualitative version of Proposition 6.2 is obtained in \([5, \text{Theorem 7}]\) which states that if \(\beta\) lies in the high temperature regime, and \(p^*\) is the unique maximizer of the variational formula \((20)\), then, for any fixed positive integer \(k\),

\[
\lim_{n \to \infty} \mathbb{P}(X_1 = a_1, \cdots, X_k = a_k) = (p^*)^{\sum_i a_i}(1 - p^*)^{k - \sum_i a_i},
\]

but does not provide any fine information about the covariance structure.

**Proof of Proposition 6.2.** Let us first consider the case when the edge \(j\) does not share a vertex with edges \(i_1, \cdots, i_k\). In order to prove \((90)\), it suffices to check that

\[
\left| \mathbb{P}(X_j = 1|X_{i_1} = a_1, \cdots, X_{i_k} = a_k) - \mathbb{P}(X_j = 1) \right| = O\left(\frac{1}{n^2}\right).
\]

Due to \((92)\), it reduces to prove that

\[
\left| \mathbb{P}(X_j = 1, X_{i_1} = a_1, \cdots, X_{i_k} = a_k) - \mathbb{P}(X_j = 1)\mathbb{P}(X_{i_1} = a_1, \cdots, X_{i_k} = a_k) \right| = O\left(\frac{1}{n^2}\right).
\]

Let \(l\) be the number of 0’s in \(a_1, \cdots, a_k\). Let us prove \((94)\) by the induction on \(l\). Suppose that \(l = 0\), in other words, \(a_1 = \cdots = a_k = 1\). Then, the LHS of \((94)\) is equal to

\[
\text{Cov}(X_j, X_{i_1} \cdots X_{i_k}).
\]

Since \((X_{i_1}, \cdots, X_{i_k}, X_j)\) satisfies the condition \((88)\) due to FKG inequality, applying \((89)\) to \(\text{Cov}(X_j, X_{i_1} \cdots X_{i_k})\) and using \((86)\),

\[
|\text{Cov}(X_j, X_{i_1} \cdots X_{i_k})| \leq C(\text{Cov}(X_j, X_{i_1}) + \cdots + \text{Cov}(X_j, X_{i_k})) = O\left(\frac{1}{n^2}\right).
\]

This concludes the proof of \((94)\) when \(l = 0\). Suppose that \((94)\) holds for \(l\), and let us prove it for \(l + 1\). Without loss of generality, assume that \(a_k = 0\). Then,

\[
\mathbb{P}(X_j = 1, X_{i_1} = a_1, \cdots, X_{i_k} = a_k) - \mathbb{P}(X_j = 1)\mathbb{P}(X_{i_1} = a_1, \cdots, X_{i_k} = a_k)
\]

\[
= \mathbb{P}(X_j = 1, X_{i_1} = a_1, \cdots, X_{i_{k-1}} = a_{k-1}) - \mathbb{P}(X_j = 1, X_{i_1} = a_1, \cdots, X_{i_{k-1}} = a_{k-1}, X_{i_k} = 1)
\]

\[
\leq C(\text{Cov}(X_j, X_{i_1}) + \cdots + \text{Cov}(X_j, X_{i_k})) = O\left(\frac{1}{n^2}\right).
\]
By the induction hypothesis, the absolute value of the above expression is $O\left(\frac{1}{n^l}\right)$. This concludes the proof of (94) for general $l$. The proof of (91) when the edge $j$ shares vertices with some of the edges $i_1, \cdots, i_k$ is same as above, except that

$$|\text{Cov}(X_j, X_{i_1 \cdots X_{i_k}})| \leq C(\text{Cov}(X_j, X_{i_1}) + \cdots + \text{Cov}(X_j, X_{i_k})) = O\left(\frac{1}{n}\right)$$

by (87) and (89).

Using the quantitative independence result Proposition 6.2 and Theorem 1, we obtain the following estimate on the $k$-correlation $\mathbb{E}[(X_{i_1} - \mathbb{E}X_{i_1}) \cdots (X_{i_k} - \mathbb{E}X_{i_k})]$.

**Proposition 6.4.** For any given integers $l, m \geq 0$ and $a_1, \cdots, a_m \geq 1$ satisfying $a_1 + \cdots + a_m \leq 2m$, there exists a constant $C_{l,a_1,\cdots,a_m} > 0$ such that the following statement holds for sufficiently large $n$: for any set of edges $i_1, \cdots, i_m, j_1, \cdots, j_l$ not sharing a vertex,

$$|\mathbb{E}[(X_{i_1} - \mathbb{E}X_{i_1})^{a_1} \cdots (X_{i_m} - \mathbb{E}X_{i_m})^{a_m} | X_{j_1}, \cdots, X_{j_l}]| \leq \frac{C_{l,a_1,\cdots,a_m}}{n^{m -(a_1 + \cdots + a_m)/2}}. \quad (95)$$

In particular, for any given positive integer $k$, for $k$ different edges $i_1, \cdots, i_k$ who do not share a vertex,

$$|\mathbb{E}[(X_{i_1} - \mathbb{E}X_{i_1}) \cdots (X_{i_k} - \mathbb{E}X_{i_k})]| = O\left(\frac{1}{n^{k/2}}\right). \quad (96)$$

**Proof.** It suffices to simply prove (95), since (96) directly follows by taking $m = k, l = 0, a_1 = a_2 \cdots = a_m = 1$. Define $\tilde{X}_i := X_i - \mathbb{E}X_i$, and let us prove (95) by the induction on $m$. Without loss of generality, we assume that $j_1 = 1, \cdots, j_l = l$, $\{i_1, \cdots, i_m\} \subset \{l+1, \cdots, [n/2]\}$, and the edges $1, \cdots, [n/2]$ do not share a vertex. The case when $m = 1$ is obvious. In fact, according to Proposition 6.2, for any fixed $l \geq 0$,

$$|\mathbb{E}[\tilde{X}_{i_1}|X_1, \cdots, X_l]| = O\left(\frac{1}{n}\right),$$

which implies (95) for $a_1 = 1$. Also, it is obvious that $|\tilde{X}_{i_1}| \leq 1$, which implies (95) in the case $a_1 = 2$.

Now suppose that (95) is true for $m \leq k - 1$, and let us prove it for $m = k$. First, consider the case when at least one of $a_1, \cdots, a_k$ is greater than one. Assuming that $a_1 \geq 2$, we have

$$\mathbb{E}[\tilde{X}_{i_1}^{a_1} \cdots \tilde{X}_{i_k}^{a_k} | X_1, \cdots, X_l] = \mathbb{E}[\tilde{X}_{i_1}^{a_1} \mathbb{E}[\tilde{X}_{i_2}^{a_2} \cdots \tilde{X}_{i_k}^{a_k} | X_1, \cdots, X_l, X_{i_1}]|X_1, \cdots, X_l]. \quad (97)$$

Since $a_2 + \cdots + a_k \leq 2(k-1)$, by the induction hypothesis,

$$|\mathbb{E}[\tilde{X}_{i_1}^{a_1} \cdots \tilde{X}_{i_k}^{a_k} | X_1, \cdots, X_l, X_{i_1}]| \leq \frac{C_{l+1,a_2,\cdots,a_k}}{n^{m-1-(a_2+\cdots+a_k)/2}} \leq \frac{C_{l+1,a_2,\cdots,a_k}}{n^{m-(a_1+\cdots+a_k)/2}}. \quad (98)$$

Since $|\tilde{X}_{i_1}| \leq 1$, (97) and (98) conclude the proof of (95). Now, let us consider the case when $a_1 = \cdots = a_k = 1$. In other words, let us prove that for any fixed $l \geq 0$,

$$|\mathbb{E}[\tilde{X}_{i_1} \cdots \tilde{X}_{i_k} | X_1, \cdots, X_l]| = O\left(\frac{1}{n^{k/2}}\right). \quad (99)$$
According to the Gaussian concentration Theorem 1,
\[ \Pr(|\tilde{X}_{i+1} + \cdots + \tilde{X}_{[n/2]}| > t) \leq 2 \exp \left(-\frac{ct^2}{n/2-1}\right) \]  
(100)

(Since the Lipschitz vector of \(X_{i+1} + \cdots + X_{[n/2]}\) is simply a vector with 1 corresponding to the edges \(l + 1\) to \([n/2]\) and 0 everywhere else, the \(L^1\) and \(L^\infty\) norms are \([n/2] - l\) and 1 respectively).

Note that due to (92), for any \(l \geq 0\), there exists a constant \(C_l > 0\) such that for sufficiently large \(n\) and any event \(A\),
\[ \Pr(A|X_1, \ldots, X_l) \leq C_l \Pr(A). \]

Combining this with (100), for any fixed integers \(l \geq 0\) and \(k \geq 1\), we have the conditional concentration result
\[ \left| \mathbb{E}\left[(\tilde{X}_{i+1} + \cdots + \tilde{X}_{[n/2]})^k|X_1, \ldots, X_l\right]\right| = O(n^{k/2}). \]  
(101)

Expanding left-hand-side of (101), we have terms like \(\mathbb{E}[\tilde{X}_{b_1}^{c_1} \cdots \tilde{X}_{b_L}^{c_L}|X_1, \ldots, X_l]\) with \(c_1 \geq \cdots \geq c_L \geq 1\) and \(c_1 + \cdots + c_L = k\). If \(c_1 \geq 2\) and \(k = c_1 + \cdots + c_L \leq 2L\), then since \(L \leq k - 1\), by the induction hypothesis,
\[ \left| \mathbb{E}[\tilde{X}_{b_1}^{c_1} \cdots \tilde{X}_{b_L}^{c_L}|X_1, \ldots, X_l]\right| \leq \frac{C_l c_1 \cdots c_L}{n^{L-(c_1+\cdots+c_L)/2}} = \frac{C_l c_1 \cdots c_L}{n^{L-k/2}}. \]

Since the number of such terms is less than \(n^L\), the sum of such terms is bounded by \(C_l c_1 \cdots c_L n^{k/2}\).

Whereas, if \(k = c_1 + \cdots + c_L \geq 2L + 1\), then each term \(\mathbb{E}[\tilde{X}_{b_1}^{c_1} \cdots \tilde{X}_{b_L}^{c_L}|X_1, \ldots, X_l]\) is bounded by 1, and thus sum of such terms is bounded by \(n^L \leq n^{(k-1)/2}\). Therefore, applying the aforementioned facts to (101), one can conclude that the absolute value of the sum of terms \(\mathbb{E}[\tilde{X}_{b_1} \cdots \tilde{X}_{b_L}|X_1, \ldots, X_l]\) is bounded by \(O(n^{k/2})\). Since there are \(\Theta(n^k)\) many such terms, we obtain (99) by symmetry which implies they have the same values.

\[ \square \]

Remark 6.5. Note that in Proposition 6.4, we only obtained higher order correlation estimates for edges not sharing a vertex. The reason behind this is as follows. Recall that FKG inequality, which ensures the non-negativity of pairwise-correlations, played a crucial role in obtaining the estimates (86) and (87) using the Gaussian concentration estimate (85) and the underlying symmetry. However, in general, one cannot ensure the positivity of higher order correlations just under a monotone measure. This is the main obstacle we face.

To elaborate on this, note first that the Gaussian concentration estimate yields the following for any \(m \leq M = \binom{n}{2}\):
\[ \mathbb{E}|\tilde{X}_1 + \cdots + \tilde{X}_m|^k \leq C m^{k/2}. \]  
(102)

Expanding the left hand side, we obtain various correlation terms of the form \(\mathbb{E}(\tilde{X}_{i_1}^{a_1} \tilde{X}_{i_2}^{a_2} \cdots \tilde{X}_{i_t}^{a_t})\). Since they are not all guaranteed to be positive, apriori we cannot rule out large correlations with different signs across different monomials canceling each other out.

Further, even terms corresponding to a particular type of a monomial are not expected to have the same value, and will depend on the graph structure induced by the edges \(i_1, i_2, i_3, \ldots, i_t\).

This forces us to work with disjoint edges so that there is a single graph structure for any monomial. This means we can take \(m\) to be at most \(O(n)\). In this case, in (102), the expansion yields \(n^k\) terms of the form \(\mathbb{E}(\tilde{X}_{i_1} \tilde{X}_{i_2} \cdots \tilde{X}_{i_k})\) and the right hand side is \(O(n^{k/2})\). This allows us, using
induction and symmetry, to prove the $O\left(\frac{1}{n^{k/2}}\right)$ bound on the $k$-point correlation. The above discussion also indicates that to improve on this, one needs to quantify the correlation depending on the graph structure induced by the corresponding edges, which is significantly more delicate.

Given the above preparation we can now finish the proof of Theorem 2.

Proof of Theorem 2. Let $S_m := \frac{X_1 + \cdots + X_m - \mathbb{E}[X_1 + \cdots + X_m]}{\sqrt{\text{Var}(X_1 + \cdots + X_m)}}$ and $\tilde{X}_i := X_i - \mathbb{E}X_i$. Throughout the proof, we fix a positive integer $k$, and let us compute the $k$-th moment of $\tilde{X}_{i_1} + \cdots + \tilde{X}_{i_m}$:

$$
\mathbb{E}[ (\tilde{X}_{i_1} + \cdots + \tilde{X}_{i_m})^k ] = \sum_{a_1, \ldots, a_m \geq 0, \ a_1 + \cdots + a_m = k} \binom{k}{a_1, \ldots, a_m} \mathbb{E}\tilde{X}_{i_1}^{a_1} \cdots \tilde{X}_{i_m}^{a_m}.
$$

For $b_1 \geq b_2 \geq \cdots \geq b_m \geq 0$ satisfying $b_1 + \cdots + b_m = k$, define $A(b_1, \ldots, b_m)$ to be a collection of $(a_1, \ldots, a_m)$s whose non-increasing rearrangement is $(b_1, \ldots, b_m)$. We claim that unless

$$
b_1 = \cdots = b_{k/2} = 2, b_{k/2+1} = \cdots = b_m = 0,
$$

$$
\sum_{(a_1, \ldots, a_m) \in A(b_1, \ldots, b_m)} \binom{k}{a_1, \ldots, a_m} \left| \mathbb{E}[\tilde{X}_{i_1}^{a_1} \cdots \tilde{X}_{i_m}^{a_m}] \right| = o(m^{k/2}). \tag{103}
$$

Note that $k/2$ is not an integer for $k$ odd, which means that (103) holds for any $b_1, \ldots, b_m$ when $k$ is odd. Let $f(b_1, \ldots, b_m)$ be the number of non-zero values in $b_1, \ldots, b_m$. According to (95), if $f(b_1, \ldots, b_m) > k/2$, then for each $(a_1, \ldots, a_m) \in A(b_1, \ldots, b_m)$,

$$
\left| \mathbb{E}[\tilde{X}_{i_1}^{a_1} \cdots \tilde{X}_{i_m}^{a_m}] \right| = O\left(\frac{1}{m^{f(b_1, \ldots, b_m)-k/2}}\right).
$$

Since $|A(b_1, \ldots, b_m)| \leq m f(b_1, \ldots, b_m)$ and $m = o(n)$, one can conclude that if $f(b_1, \ldots, b_m) > k/2$, then

$$
\sum_{(a_1, \ldots, a_m) \in A(b_1, \ldots, b_m)} \left| \mathbb{E}[\tilde{X}_{i_1}^{a_1} \cdots \tilde{X}_{i_m}^{a_m}] \right| = O\left(\frac{m f(b_1, \ldots, b_m)}{n^{f(b_1, \ldots, b_m)-k/2}}\right) = o(m^{k/2}). \tag{104}
$$

Since $|\mathbb{E}[\tilde{X}_{i_1}^{a_1} \cdots \tilde{X}_{i_m}^{a_m}]| \leq 1$, if $f(b_1, \ldots, b_m) < k/2$, then

$$
\sum_{(a_1, \ldots, a_m) \in A(b_1, \ldots, b_m)} \left| \mathbb{E}[\tilde{X}_{i_1}^{a_1} \cdots \tilde{X}_{i_m}^{a_m}] \right| = O(m f(b_1, \ldots, b_m)) = o(m^{k/2}). \tag{105}
$$

In the case when $k$ is even and $f(b_1, \ldots, b_m) = k/2$, we prove that unless $b_1 = \cdots = b_{k/2} = 2$ and $b_{k/2+1} = \cdots = b_m = 0$, for $(a_1, \ldots, a_m) \in A(b_1, \ldots, b_m)$,

$$
\left| \mathbb{E}[\tilde{X}_{i_1}^{a_1} \cdots \tilde{X}_{i_m}^{a_m}] \right| = O\left(\frac{1}{m^{1/2}}\right). \tag{106}
$$

We have $b_1 \geq 3$ unless $b_1 = \cdots = b_{k/2} = 2$ and $b_{k/2+1} = \cdots = b_m = 0$. Without loss of generality, assume that $a_j = b_j$ for $j = 1, 2, \ldots, m$. Note that

$$
\mathbb{E}[\tilde{X}_{i_1}^{a_1} \cdots \tilde{X}_{i_m}^{a_m}] = \mathbb{E}[\tilde{X}_{i_1}^{a_1} \mathbb{E}[\tilde{X}_{i_2}^{a_2} \cdots \tilde{X}_{i_m}^{a_m} | X_{i_1}]]. \tag{107}
$$

Since $a_1 + \cdots + a_m = k$ and $a_1 \geq 3$, we have $a_2 + \cdots + a_m \leq k - 3$. Therefore, by (95), we have

$$
\left| \mathbb{E}[\tilde{X}_{i_2}^{a_2} \cdots \tilde{X}_{i_m}^{a_m} | X_{i_1}] \right| = O\left(\frac{1}{n^{k/2-1-(a_2+\cdots+a_m)/2}}\right) = O\left(\frac{1}{n^{1/2}}\right). \tag{108}
$$
Therefore, (107) and (108) concludes the proof of (106). Using the fact $|A(b_1, \ldots, b_m)| \leq m^{f(b_1, \ldots, b_m)}$, (106) implies that when $k$ is even and $f(b_1, \ldots, b_m) = k/2$, unless $b_1 = \cdots = b_{k/2} = 2$ and $b_{k/2 + 1} = \cdots = b_m = 0,$

$$\sum_{(a_1, \ldots, a_m) \in A(b_1, \ldots, b_m)} \mathbb{E}[\tilde{X}_{i_1}^{a_1} \cdots \tilde{X}_{i_m}^{a_m}] = O\left(\frac{m^{k/2}}{n^{1/2}}\right) = o(m^{k/2}).$$

(109)

Since the multinomial coefficient $\binom{k}{a_1, \ldots, a_m}$ is always bounded by $k!$, we obtain (103) by (104), (105), and (109).

Using (103), one concludes that if $k$ is odd, then

$$\left|\mathbb{E}[\tilde{X}_{i_1} + \cdots + \tilde{X}_{i_m}]^k\right| = o(m^{k/2}).$$

(110)

Whereas, if $k$ is even, then

$$\mathbb{E}[\tilde{X}_{i_1} + \cdots + \tilde{X}_{i_m}]^k = o(m^{k/2}) + \binom{k}{2, \ldots, 2} \sum_{j_1 < \cdots < j_{k/2}} \mathbb{E}\tilde{X}_{i_{j_1}}^2 \cdots \tilde{X}_{i_{j_{k/2}}}^2.$$

(111)

Note that by symmetry,

$$\text{Var}(X_{i_1} + \cdots + X_{i_m}) = m\text{Var}(X_{i_1}) + m(m - 1)\text{Cov}(X_{i_1}, X_{i_2}).$$

(112)

Since the law of $X_{i_1}$ converges weakly to the Bernoulli distribution with a parameter $p^*$,

$$\lim_{n \to \infty} \text{Var}(X_{i_1}) = p^*(1 - p^*).$$

Recall that Cov$(X_{i_1}, X_{i_2}) = O\left(\frac{1}{n^2}\right)$ since $i_1$ and $i_2$ do not share a vertex (see (86)). Thus, applying this to (112),

$$\lim_{n \to \infty} \frac{1}{m} \text{Var}(X_{i_1} + \cdots + X_{i_m}) = p^*(1 - p^*).$$

(113)

Therefore, by (110) and (113), if $k$ is odd, then

$$\lim_{n \to \infty} \text{E}S_m^k = 0.$$  

(114)

Whereas, if $k$ is even, then by (111) and symmetry,

$$\text{E}S_m^k = o(1) + \binom{k}{2, \ldots, 2}\binom{m}{k/2} \frac{\mathbb{E}\tilde{X}_{i_{j_1}}^2 \cdots \tilde{X}_{i_{j_{k/2}}}^2}{\text{Var}(X_{i_1} + \cdots + X_{i_m})^{k/2}}.$$  

(115)

According to the asymptotic independence result (92), we have

$$\lim_{n \to \infty} \mathbb{E}[\tilde{X}_{i_{j_1}}^2 \cdots \tilde{X}_{i_{j_{k/2}}}^2] = (p^*(1 - p^*))^{k/2}.$$  

(116)

Therefore, applying (113) and (116) to (115), one can conclude that for $k$ even,

$$\lim_{n \to \infty} \text{E}S_m^k = (k - 1)!!.$$  

(117)

Here, $(k - 1)!!$ denotes the product of all numbers from $1$ to $k - 1$ that have the same parity as $k - 1$. Thus, (114) and (117) imply that the normalized sum $S_m$ weakly converges to the standard normal distribution, by, say, the classical Fréchet-Shohat theorem [25] along with the fact that the normal distribution is a determinate measure, i.e., its moments determine it uniquely. \qed
7. Bounding $W_1$-Wasserstein distance: Proof of Theorem 3

We start with the following related result about one point marginals.

**Proposition 7.1.** For any edge $i$, 

$$|\mathbb{E}X_i - p^*| = O\left(\frac{\sqrt{\log n}}{\sqrt{n}}\right).$$  \hspace{1cm} (118)  

**Proof.** By symmetry, let us consider the case $i = 1$. Note that 

$$\mathbb{E}X_1 = \mathbb{E}[\mathbb{E}[X_1|X_2,\cdots,X_M]] = \mathbb{E}\Phi(\partial_1 H(X)),$$  \hspace{1cm} (119)  

where in the last inequality we used 

$$\mathbb{E}[X_1|X_2,\cdots,X_M] = \mathbb{P}(X_1 = 1|X_2,\cdots,X_M) = \Phi(\partial_1 H(X)).$$

The Lipschitz constant of $\partial_1 H(x)$ with respect to the edge $j$ is $O\left(\frac{1}{n}\right)$ if $j$ does not share a vertex with the edge 1 and $O\left(\frac{1}{n}\right)$ otherwise. To see this, note first that by (37), 

$$\partial_1 H(x) = \sum_{i=1}^{s} \beta_i \frac{N_{G_i}(x,1)}{n|V_i| - 2}.$$  

As mentioned in the proof of Proposition 4.3, the quantity $N_{G_i}(x_{f+},1) - N_{G_i}(x_{f-},1) = N_{G_i}(x,1,f)$ is $O(n|V_i|^{-4})$ if $f$ does not share a vertex with the edge 1 and $O(n|V_i|^{-3})$ if $f$ shares a vertex with the edge 1. Due to the normalization factor $n|V_i|^{-2}$, we have the asserted bound on the entries of the Lipschitz vector for $\partial_1 H(x)$.

Hence, its $L^1$ and $L^\infty$ norms are $O(1)$ and $O\left(\frac{1}{n}\right)$ respectively. Thus, according to Theorem 1, for some constant $c > 0$, for all $\varepsilon > 0$ 

$$\mathbb{P}(|\partial_1 H(X) - \mathbb{E}\partial_1 H(X)| > \varepsilon) \leq 2\exp(-cn\varepsilon^2)$$  \hspace{1cm} (120)  

For $K > 0$ chosen later, let us denote 

$$A_n := \mathbb{E}\left[\Phi(\partial_1 H(X)) : |\partial_1 H(X) - \mathbb{E}\partial_1 H(X)| \leq \frac{K}{\sqrt{n}}\right],$$  

$$B_n := \mathbb{E}\left[\Phi(\partial_1 H(X)) : |\partial_1 H(X) - \mathbb{E}\partial_1 H(X)| > \frac{K}{\sqrt{n}}\right].$$

Then, by (119), we have 

$$\mathbb{E}X_1 = A_n + B_n.$$  \hspace{1cm} (121)  

Since $\Phi(x) = \frac{e^x}{1+e^x}$ is non-negative and increasing, 

$$\Phi\left(\mathbb{E}\partial_1 H(X) - \frac{K}{\sqrt{n}}\right) \mathbb{P}\left(|\partial_1 H(X) - \mathbb{E}\partial_1 H(X)| \leq \frac{K}{\sqrt{n}}\right) \leq A_n \leq \Phi\left(\mathbb{E}\partial_1 H(X) + \frac{K}{\sqrt{n}}\right).$$  \hspace{1cm} (122)  

Choosing $\varepsilon = \frac{K}{\sqrt{n}}$ in (120), and using the fact that $\Phi(x) \leq 1$, 

$$\Phi\left(\mathbb{E}\partial_1 H(X) - \frac{K}{\sqrt{n}}\right) - 2\exp(-cK^2) \leq A_n \leq \Phi\left(\mathbb{E}\partial_1 H(X) + \frac{K}{\sqrt{n}}\right).$$  \hspace{1cm} (123)
Next, using $\Phi(x) \leq 1$ and (120) again, we obtain:

$$0 \leq B_n \leq \mathbb{P}\left( |\partial_1 H(X) - \mathbb{E}\partial_1 H(X)| > \frac{K}{\sqrt{n}} \right) \leq 2 \exp(-cK^2). \tag{124}$$

Now letting $\mathbb{E}X_1 = a_n$, we express $\mathbb{E}\partial_1 H(X)$ in terms of $a_n$. By (37),

$$\mathbb{E}\partial_1 H(X) = \sum_{i=1}^{s} \beta_i \mathbb{E}N_{G_i}(X, 1) \left/ \frac{n}{|l|} \right. ,$$

and note that by (5), $N_{G_i}(X, 1)$ is a sum of $2|E_i|(1+o(1))n|l|^{-2}$ many terms of the type $X_{l_1}X_{l_2} \cdots X_{l_{|E_i|-1}}$, with distinct edges $l_1, \ldots, l_{|E_i|-1}$. Thus, using the definition of $\Psi_\beta$ from (17), by (84), for some constant $C > 0$

$$\left| \mathbb{E}\partial_1 H(X) - \Psi_\beta(a_n) \right| \leq \frac{C}{n}. \tag{125}$$

Therefore, applying (123), (124), and (125) to (121),

$$\Phi\left( \Psi_\beta(a_n) - \frac{K}{\sqrt{n}} - \frac{C}{n} \right) - 2 \exp(-cK^2) \leq a_n \leq \Phi\left( \Psi_\beta(a_n) + \frac{K}{\sqrt{n}} + \frac{C}{n} \right) + 2 \exp(-cK^2).$$

Since one can check that $\sup_{x \in \mathbb{R}} |\Phi'(x)| \leq \frac{1}{4}$,

$$-\frac{1}{4} \left( \frac{K}{\sqrt{n}} + \frac{C}{n} \right) - 2 \exp(-cK^2) \leq a_n - \Phi(\Psi_\beta(a_n)) \leq \frac{1}{4} \left( \frac{K}{\sqrt{n}} + \frac{C}{n} \right) + 2 \exp(-cK^2). \tag{126}$$

Recall that $p^*$ satisfies the fixed point equation

$$p^* = \Phi(\Psi_\beta(p^*)) = \frac{\exp(\Psi_\beta(p^*))}{1 + \exp(\Psi_\beta(p^*))}. \tag{127}$$

Denoting by $f$ the map $t \mapsto t - \Phi(\Psi_\beta(t))$, note that $f(0) < 0$ and $f(1) > 0$. Further, since $\beta$ is in the high temperature regime, the map $f$ has a unique root $p^*$. Also recall from (19) that $f'(p^*) > d$ for some positive constant $d$. Thus, owing to continuity of $f$ on the compact interval $[0, 1]$, it follows that for all small enough $\varepsilon > 0$, for all $t \in [0, 1]$ with $|t - p^*| \geq \varepsilon$, we have $|f(t)| > \frac{d}{2} \varepsilon$.

Hence, by (126), for all $K$ large enough,

$$|p^* - a_n| \leq \frac{2}{d} \left( \frac{1}{4} \left( \frac{K}{\sqrt{n}} + \frac{C}{n} \right) + 2 \exp(-cK^2) \right).$$

Taking $K = C' \sqrt{\log n}$ with a sufficiently large constant $C' > 0$ in the above inequality, we obtain (118). \hfill \Box

Using the above, we now finish the proof of Theorem 3.

Proof of Theorem 3. Suppose that two random variables $X$ and $Y$ are distributed as $\text{ERGM}(\beta)$ and $G(n, p^*)$, respectively. In order to prove (27), it suffices to construct a coupling of $(X, Y)$ such that

$$\mathbb{E}(d_H(X, Y)) \leq Cn^{3/2} \sqrt{\log n}. \tag{128}$$

Consider two different GD $X(t)$ and $Y(t)$ associated with $\text{ERGM}(\beta)$ and $G(n, p^*)$, respectively. Assume that $X(t)$ and $Y(t)$ start from initial distributions $\pi = \text{ERGM}(\beta)$ and $\nu = G(n, p^*)$ so that each GD is stationary. Let us couple the initial state $X(0)$ and $Y(0)$ in an arbitrary way. We then couple $X(t)$ and $Y(t)$ inductively in the following natural way: assume that $X(t)$-chain is at $x$ and
Thus, using (by replacing \( x_I \) by \( Z_1^I \) and similarly \( Y(t + 1) \) is the same as \( y \) with \( y_I \) replaced by \( Z_2^I \). Now note that given \( X(t) \) and \( Y(t) \), we have,

\[
\mathbb{P}(X_i(t + 1) \neq Y_i(t + 1)) \leq \left( 1 - \frac{1}{M} \right) \mathbb{I}\{X_i(t) \neq Y_i(t)\} + \frac{1}{M} d_{TV}(\pi_i(\cdot|X(t)), \nu_i(\cdot|Y(t))).
\]  

(129)

Since \( p^* = \Phi(\Psi_B(p^*)) \), by the mean value theorem and that \( \sup_{x \in \mathbb{R}} |\Phi'(x)| \leq \frac{1}{4} \),

\[
d_{TV}(\pi_i(\cdot|X(t)), \nu_i(\cdot|Y(t))) = |\Phi(\partial_i H(X(t))) - p^*| \leq \frac{1}{4} |\partial_i H(X(t)) - \Psi_B(p^*)|.
\]

(130)

Since \( X(t) \) is stationary, by (125),

\[
\left| \mathbb{E}(\partial_i H(X(t)) - \sum_{j=1}^{s} 2\beta_j |E_j| (\mathbb{E} X_1)^{|E_j| - 1} \right| = O\left( \frac{\sqrt{\log n}}{\sqrt{n}} \right).
\]

Using this with Proposition 7.1, we get,

\[
|\mathbb{E}(\partial_i H(X(t)) - \Psi_B(p^*))| = O\left( \frac{\sqrt{\log n}}{\sqrt{n}} \right).
\]

(131)

Since \( X(t) \) is distributed as ERGM(\( \beta \)), recalling (120),

\[
\mathbb{P}\left( |\partial_i H(X(t)) - \mathbb{E}(\partial_i H(X(t)))| > C_0 \frac{\sqrt{\log n}}{\sqrt{n}} \right) \leq 2e^{-c n(C^2 \log n/n)} = 2n^{-cC^2}.
\]

(132)

By (131) and (132), for all large constant \( C_0 > 0 \), there exists \( C_1 > cC^2/2 \) such that

\[
\mathbb{P}\left( |\partial_i H(X(t)) - \Psi_B(p^*)| > C_0 \frac{\sqrt{\log n}}{\sqrt{n}} \right) \leq n^{-C_1}.
\]

(133)

Let us assume that \( C_0 \) is a sufficiently large constant so that \( C_1 > 3 \). By (129) and (130), given \( X(t) \) and \( Y(t) \), on the event \( |\partial_i H(X(t)) - \Psi_B(p^*)| \leq C_0 \frac{\sqrt{\log n}}{\sqrt{n}} \),

\[
\mathbb{P}(X_i(t + 1) \neq Y_i(t + 1)) \leq \left( 1 - \frac{1}{M} \right) \mathbb{I}(X_i(t) \neq Y_i(t)) + \frac{C_0 \sqrt{\log n}}{4M} \sqrt{n}.
\]

(134)

Whereas, on the event \( |\partial_i H(X(t)) - \Psi_B(p^*)| > C_0 \frac{\sqrt{\log n}}{\sqrt{n}} \), we use the trivial bound

\[
\mathbb{P}(X_i(t + 1) \neq Y_i(t + 1)) \leq 1.
\]

(135)

Thus, using (133), (134), and (135),

\[
\mathbb{P}(X_i(t + 1) \neq Y_i(t + 1)) \leq \left( 1 - \frac{1}{M} \right) \mathbb{P}(X_i(t) \neq Y_i(t)) + \frac{C_0 \sqrt{\log n}}{4M} \sqrt{n} + n^{-C_1}.
\]

Recalling that \( M = \frac{n(n-1)}{2} \) and \( C_1 > 3 \), there exists a constant \( C > 0 \) such that for sufficiently large \( n \), for \( t \geq c'n^{5/2} \) with \( c' > 0 \), (any \( t \) suitably greater than \( n^2 \) will do),

\[
\mathbb{P}(X_i(t) \neq Y_i(t)) \leq C \frac{\sqrt{\log n}}{\sqrt{n}}.
\]

Adding this over all edges \( i \), we obtain \( \mathbb{E}(d_H(X(t), Y(t))) \leq C n^{3/2} \sqrt{\log n} \). Since \( X(t) \) and \( Y(t) \) are distributed as ERGM(\( \beta \)) and \( G(n, p^*) \) respectively, we obtain (27). \( \square \)
Remark 7.2. As the reader might notice, in the above proof, instead of choosing edges uniformly to update, one can also perform a sequential update of all the edges. More precisely, ordering the edges in an arbitrary fashion labelling them $1, \ldots, M$, let $(X(t))_{t=0,1,\ldots,M}$ be a Markov chain starting from the initial distribution $X(0) = \text{ERGM}(\beta)$ such that given $X(t-1)$, $X(t)$ is obtained by resampling the $t$th edge, according to the conditional distribution given other edges. Similarly, define the corresponding Markov chain $Y(t)_{t=0,1,\ldots,M}$ for $G(n,p^*)$. It is obvious that both $X(t)$ and $Y(t)$ are stationary. Now consider the natural coupling of the two Markov chains by taking an arbitrary coupling $(X(0), Y(0))$ and at each time $t \geq 1$, given $(X(t-1), Y(t-1))$, taking an optimal coupling for $X_t(t)$ and $Y_t(t)$ that attains the total variation distance. Then, by (130) and (133),

$$\mathbb{P}(X_{t+1}(t+1) \neq Y_{t+1}(t+1)) = d_{\text{TV}}(\pi_t(\cdot | X(t)), \nu_t(\cdot | Y(t))) \leq \frac{C_0}{\sqrt{n}} \log n + n^{-C_1}. $$

Adding this over $t = 1, 2, \ldots, M$, since $C_1 > 3$, we have

$$\mathbb{E}(d_H(X(M), Y(M))) = \sum_{t=1}^{M} \mathbb{P}(X_t(M) \neq Y_t(M)) = \sum_{t=1}^{M} \mathbb{P}(X_t(t) \neq Y_t(t)) \leq M \left( \frac{C_0}{\sqrt{n}} \log n + n^{-C_1} \right) = O(n^{3/2} \sqrt{\log n}).$$

Since two chains $X(t)$ and $Y(t)$ are stationary, this concludes the proof.

7.1. Pseudo-likelihood estimators. Being a central object in the study of statistical models on networks, a natural question about the ERGM is whether one can deduce any meaningful estimate of the parameter $\beta$ given a single realization of the graph. A well known estimator in such contexts is the so-called Maximum Likelihood Estimator (MLE), which as the name suggests, are the values of the parameter that maximizes the probability of the realized sample. However often in high dimensional models such as the ERGM, the optimization problem associated to finding the MLE, neither has a closed form solution, nor is computationally feasible. In such settings a related proxy is the notion of a pseudo-likelihood estimator. This was was first introduced by Besag [3, 4] in the context of analyzing data possessing spatial dependence. To define this precisely, consider a random vector $X = (X_1, \ldots, X_n)$ with probability density function $f(\beta, X)$, parametrized by a parameter $\beta$. Define $f_i(\beta, x)$ to be a conditional probability density of $X_i$ given $(X_j)_{j \neq i}$. The maximum pseudo-likelihood estimator (MPLE) of $\beta$ is defined by

$$\hat{\beta}_{\text{MPLE}} := \arg \max_{\beta} \prod_{i=1}^{n} f_i(\beta, X).$$

In practice, this is often much simpler to analyze than the MLE problem,

$$\hat{\beta}_{\text{MLE}} := \arg \max_{\beta} f(\beta, X).$$

A sequence $\{\hat{\beta}_n\}_n$ is said to be a consistent sequence of estimators for $\beta$ provided that for any $\varepsilon > 0$,

$$\lim_{n \to \infty} \mathbb{P}(|\hat{\beta}_n - \beta| > \varepsilon) = 0.$$
While studying the consistency of MLE in rather general settings has been an important classical theme in probability theory and statistics, more recently, there has been a parallel interest in understanding the MLEs for various statistical mechanics models, in particular for spin systems. One of the most notable results was proved by Chatterjee [9] where among various things, it was proved that the MLE in the Sherrington-Kirkpatrick (S-K) model and the Hopfield model with a single parameter $\beta > 0$ is consistent for $\beta$, whereas, consistency does not hold in the high temperature Curie-Weiss model. In fact, if the inverse temperature $\beta$ satisfies $0 \leq \beta < 1$, then $\hat{\beta}_n \to 1$ in probability (see [9, Section 1.7] for details). Later, similar problems for the Ising model on regular graphs and on Erdős-Rényi graphs were studied in [1] and more recently in [26].

While such a study for the ERGM has not yet appeared in the literature, the consistency of the MLE for the ERGM with a single parameter $\beta$ in the whole positive temperature regime $\beta > 0$ is expected to follow using the robust arguments developed in [9]. On the other hand, it seems that a general theory of the consistency of MLE with several parameters has not been developed yet, except in some special cases (see [26, 16]). In fact, the problem seems to be ill-posed in the case of the ERGM, since, in the high temperature case, using the concentration result Theorem 1, with high probability, the critical equation corresponding to finding $\hat{\beta}$, the MLE for $\beta$, seems to only say that $p^* \approx \Psi_{\hat{\beta}}(p^*)$, and hence does not provide any further information about the vector $\beta$. This might lead one to wonder for any parameters $\beta$ and $\gamma$ with $\varphi_{\beta}(p^*) = p^* = \varphi_{\gamma}(p^*)$, since both ERGM($\beta$) and ERGM($\gamma$) behave like $G(n,p^*)$, if they are contiguous. Nonetheless, while in light of the above, the MLE is unable to estimate the parameters, as was mentioned earlier, in [39] a consistent estimator was obtained for the parameters in the special case of the two-star model, ruling out the possibility of such a contiguity result, at least in this case.

**References**

[1] Anirban Basak and Sumit Mukherjee. Universality of the mean-field for the Potts model. *Probability Theory and Related Fields*, 168(3-4):557–600, 2017.
[2] Noam Berger, Claire Kenyon, Elchanan Mossel, and Yuval Peres. Glauber dynamics on trees and hyperbolic graphs. *Probability Theory and Related Fields*, 131(3):311–340, 2005.
[3] Julian Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):192–225, 1974.
[4] Julian Besag. Statistical analysis of non-lattice data. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 24(3):179–195, 1975.
[5] Shankar Bhamidi, Guy Bresler, and Allan Sly. Mixing time of exponential random graphs. *The Annals of Applied Probability*, pages 2146–2170, 2011.
[6] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.
[7] Filippo Cesi. Quasi-factorization of the entropy and logarithmic Sobolev inequalities for Gibbs random fields. *Probability Theory and Related Fields*, 120(4):569–584, 2001.
[8] Sourav Chatterjee. Concentration inequalities with exchangeable pairs (ph. d. thesis). *arXiv preprint math/0507526*, 2005.
[9] Sourav Chatterjee. Estimation in spin glasses: A first step. *The Annals of Statistics*, 35(5):1931–1946, 2007.
[10] Sourav Chatterjee. Stein’s method for concentration inequalities. *Probability Theory and Related Fields*, 138(1):305–321, 2007.
[11] Sourav Chatterjee. An introduction to large deviations for random graphs. *Bulletin of the American Mathematical Society*, 53(4):617–642, 2016.
[12] Sourav Chatterjee and Amir Dembo. Nonlinear large deviations. *Advances in Mathematics*, 299:396–450, 2016.
[13] Sourav Chatterjee and Partha S Dey. Applications of Stein’s method for concentration inequalities. *The Annals of Probability*, 38(6):2443–2485, 2010.

[14] Sourav Chatterjee and Persi Diaconis. Estimating and understanding exponential random graph models. *The Annals of Statistics*, 41(5):2428–2461, 2013.

[15] Sourav Chatterjee and SR Srinivasa Varadhan. The large deviation principle for the Erdős-Rényi random graph. *European Journal of Combinatorics*, 32(7):1000–1017, 2011.

[16] Constantinos Daskalakis, Nishanth Dikkala, and Ioannis Panageas. Regression from dependent observations. *arXiv preprint arXiv:1903.08353*, 2019.

[17] Ryan DeMuse, Terry Easlick, and Mei Yin. Mixing time of vertex-weighted exponential random graphs. *Journal of Computational and Applied Mathematics*, 362:443–459, 2019.

[18] Roland L Dobrushin. The description of a random field by means of conditional probabilities and conditions of its regularity. *Theory of Probability & Its Applications*, 13(2):197–224, 1968.

[19] Roland L Dobrushin. Prescribing a system of random variables by conditional distributions. *Theory of Probability & Its Applications*, 15(3):458–486, 1970.

[20] Ronen Eldan. Gaussian-width gradient complexity, reverse log-Sobolev inequalities and nonlinear large deviations. *Geometric and Functional Analysis*, 28(6):1548–1596, 2018.

[21] Ronen Eldan and Renan Gross. Exponential random graphs behave like mixtures of stochastic block models. *The Annals of Applied Probability*, 28(6):3698–3735, 2018.

[22] Richard S Ellis. *Entropy, large deviations, and statistical mechanics*. Springer, 2007.

[23] Richard S Ellis and Charles M Newman. The statistics of Curie-Weiss models. *Journal of Statistical Physics*, 19(2):149–161, 1978.

[24] Ove Frank and David Strauss. Markov graphs. *Journal of the American Statistical Association*, 81(395):33–50, 1986.

[25] Maurice Fréchet and James Shohat. A proof of the generalized second-limit theorem in the theory of probability. *Transactions of the American Mathematical Society*, 33(2):533–543, 1931.

[26] Promit Ghosal and Sumit Mukherjee. Joint estimation of parameters in Ising model. *arXiv preprint arXiv:1801.06570*, 2018.

[27] Larry Goldstein and Nathakhun Wiroonsri. Stein’s method for positively associated random variables with applications to the ising and voter models, bond percolation, and contact process. In *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, volume 54, pages 385–421. Institut Henri Poincaré, 2018.

[28] Paul W Holland and Samuel Leinhardt. An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, 76(373):33–50, 1981.

[29] John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558, 1982.

[30] Richard Kenyon and Mei Yin. On the asymptotics of constrained exponential random graphs. *Journal of Applied Probability*, 54(1):165–180, 2017.

[31] Christof Külske. Concentration inequalities for functions of Gibbs fields with application to diffraction and random Gibbs measures. *Communications in Mathematical Physics*, 239(1-2):29–51, 2003.

[32] Michel Ledoux. *The concentration of measure phenomenon*. Number 89. American Mathematical Soc., 2001.

[33] Michel Ledoux. Spectral gap, logarithmic Sobolev constant, and geometric bounds. *Surveys in Differential Geometry*, 9(1):219–240, 2004.

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

[35] László Lovász. *Large networks and graph limits*, volume 60. American Mathematical Soc., 2012.

[36] Fabio Martinelli and Enzo Olivieri. Approach to equilibrium of Glauber dynamics in the one phase region. *Communications in Mathematical Physics*, 161(3):447–486, 1994.

[37] Fabio Martinelli and Enzo Olivieri. Approach to equilibrium of Glauber dynamics in the one phase region. ii. the general case. *Communications in Mathematical Physics*, 161(3):487–514, 1994.

[38] Katalin Marton. Logarithmic sobolev inequalities in discrete product spaces: a proof by a transportation cost distance. *arXiv preprint arXiv:1507.02803*, 2015.
[39] Sumit Mukherjee and Yuanzhe Xu. Statistics of the two-star ergm. arXiv preprint arXiv:1310.4526, 2013.

[40] Charles M Newman. Normal fluctuations and the FKG inequalities. Communications in Mathematical Physics, 74(2):119–128, 1980.

[41] Charles M Newman. A general central limit theorem for FKG systems. Communications in Mathematical Physics, 91(1):75–80, 1983.

[42] Juyong Park and Mark EJ Newman. Solution of the two-star model of a network. Physical Review E, 70(6):066146, 2004.

[43] Juyong Park and Mark EJ Newman. Solution for the properties of a clustered network. Physical Review E, 72(2):026136, 2005.

[44] Charles Radin and Mei Yin. Phase transitions in exponential random graphs. The Annals of Applied Probability, pages 2458–2471, 2013.

[45] Gesine Reinert and Nathan Ross. Approximating stationary distributions of fast mixing Glauber dynamics, with applications to exponential random graphs. arXiv preprint arXiv:1712.05736, 2017.

[46] Holger Sambale and Arthur Sinulis. Logarithmic sobolev inequalities for finite spin systems and applications. arXiv preprint arXiv:1807.07765, 2018.

[47] Cosma Rohilla Shalizi and Alessandro Rinaldo. Consistency under sampling of exponential random graph models. Annals of Statistics, 41(2):508, 2013.

[48] David Sherrington and Scott Kirkpatrick. Solvable model of a spin-glass. Physical Review Letters, 35(26):1792, 1975.

[49] Daniel W Stroock and Boguslaw Zegarlinski. The equivalence of the logarithmic Sobolev inequality and the Dobrushin-Shlosman mixing condition. Communications in Mathematical Physics, 144(2):303–323, 1992.

[50] Daniel W Stroock and Boguslaw Zegarlinski. The logarithmic Sobolev inequality for discrete spin systems on a lattice. Communications in Mathematical Physics, 149(1):175–193, 1992.

[51] Mei Yin and Lingjiong Zhu. Asymptotics for sparse exponential random graph models. Brazilian Journal of Probability and Statistics, 31(2):394–412, 2017.

[52] Boguslaw Zegarlinski. Dobrushin uniqueness theorem and logarithmic Sobolev inequalities. Journal of Functional Analysis, 105(1):77–111, 1992.

Department of Statistics, Evans Hall, University of California, Berkeley, CA 94720, USA
Email address: sganguly@berkeley.edu

Department of Mathematics, University of California, Los Angeles, CA 90095, USA
Email address: ksnam@math.ucla.edu