APPROXIMATING STATIONARY DISTRIBUTIONS OF FAST MIXING GLAUBER DYNAMICS, WITH APPLICATIONS TO EXPONENTIAL RANDOM GRAPHS

BY GESINE REINERT[1] AND NATHAN ROSS[2]

University of Oxford and University of Melbourne

We provide a general bound on the Wasserstein distance between two arbitrary distributions of sequences of Bernoulli random variables. The bound is in terms of a mixing quantity for the Glauber dynamics of one of the sequences, and a simple expectation of the other. The result is applied to estimate, with explicit error, expectations of functions of random vectors for some Ising models and exponential random graphs in "high temperature" regimes.

1. Introduction. A high-level heuristic in statistical physics models is that systems of dependent random variables having the stationary distribution of a fast mixing Markov chain are approximately independent. Precise statements of this heuristic include spatial-temporal mixing conditions (see, e.g., Mossel and Sly (2013) and discussion there); as well as establishing properties of the dependent variables common to independent variables, such as central limit theorems (see, e.g., Ellis and Newman (1978)); and concentration inequalities (see, e.g., Luczak (2008), Chatterjee ((2005), Chapter 4), and Gheissari, Lubetzky and Peres (2018)). Such results are particularly useful for models with intractable normalizing constants; for example, Ising models and exponential random graphs; since it is possible to define and analyse Markov chains which only depend on ratios of probabilities and, therefore, do not require computation of the normalizing constant.

We provide a general result (Theorem 2.1 below) that explicitly bounds the Wasserstein distance between distributions of vectors of Bernoulli random variables, in terms of (1) a quantity depending on the mixing of the Glauber dynamics of one of the vectors; and (2) a straightforward expectation of the other. The result applies to many models where fast mixing is known, and here we apply it to Ising models and exponential random graphs, sharpening some recent results of Eldan and Gross (2018a, 2018b).

We now state a simplified version of our general approximation result, and its implications in some applications; (1.7) and Propositions 1.11 and 1.12. Let $N \in$

$\mathbb{N}$ and $X = (X_1, \ldots, X_N) \in \{0, 1\}^N$ be a vector of Bernoulli random variables. Denote $[N] := \{1, \ldots, N\}$, and for $s \in [N]$, write $e_s \in \{0, 1\}^N$ for the vector with 1 in the sth coordinate, and 0 in all others, and, for $x \in \{0, 1\}^N$, denote

$$x^{(s,1)} = x + (1 - x_s)e_s, \qquad x^{(s,0)} = x - x_s e_s,$$

so that $x^{(s,1)}$ has a 1 in the sth coordinate and is otherwise the same as x, and similar for $x^{(s,0)}$ except there is a 0 in the sth coordinate.

Define the Glauber dynamics Markov chain $\{X(m), m = 0, 1, \ldots\}$ with transition probabilities

$$\mathbb{P}\big(x \mapsto x^{(s,1)}\big) = \frac{1}{N} - \mathbb{P}\big(x \mapsto x^{(s,0)}\big) = \frac{q_X(x^{(s,1)}|x)}{N}, \quad s \in [N],$$

where

$$q_X\big(x^{(s,1)}|x\big) := \mathbb{P}\big(X_s = 1|(X_u)_{u \neq s} = (x_u)_{u \neq s}\big).$$

In words, at every step in the chain, we choose a coordinate uniformly at random and resample it conditional on the values at the other coordinates. It is easily checked that this chain is reversible with respect to $\mathscr{L}(X)$.

For functions $h : \{0, 1\}^N \to \mathbb{R}$, define

$$\Delta_s h(x) := h\big(x^{(s,1)}\big) - h\big(x^{(s,0)}\big),$$

denote the supremum norm by $\| \cdot \|$, let $\|\Delta h\| := \sup_{s \in [N]} \|\Delta_s h\|$, and denote the Hamming distance of $x, y \in \{0, 1\}^N$ by $d_{\mathrm{H}}(x, y) := \sum_{s=1}^N |x_s - y_s|$.

THEOREM 1.1. *Let $X, Y \in \{0, 1\}^N$ be random vectors. Write*

$$(1.1) \qquad \mathbb{P}(X = x) = \frac{1}{\kappa} \exp\{L(x)\},$$

for a function L and normalising constant κ, and assume that Y has independent coordinates with $p_s := \mathbb{P}(Y_s = 1)$, for $s = 1, \ldots, n$, satisfying

$$(1.2) \qquad p_s = \frac{1 + \tanh(\frac{1}{2}\mathbb{E}\Delta_s L(Y))}{2}.$$

Assume that for each $s \in [N]$ and $x \in \{0, 1\}^N$, there is a coupling $(\tilde{U}^{[x,s]}, \tilde{V}^{[x,s]})$ with $\mathcal{L}(\tilde{U}^{[x,s]}) = \mathcal{L}(X(1)|X(0) = x^{(s,1)})$ and $\mathcal{L}(\tilde{V}^{[x,s]}) = \mathcal{L}(X(1)|X(0) = x^{(s,0)})$, such that for some $0 < \rho \leq 1$,

$$(1.3) \qquad \mathbb{E}\, d_{\mathrm{H}}\big(\tilde{U}^{[x,s]}, \tilde{V}^{[x,s]}\big) \leq (1 - \rho),$$

and that the Glauber dynamics of $\mathscr{L}(X)$ are irreducible. Then for any $h : \{0, 1\}^N \to \mathbb{R}$,

$$(1.4) \qquad \big|\mathbb{E}h(X) - \mathbb{E}h(Y)\big| \leq \frac{\|\Delta h\|}{4N\rho} \sum_{s=1}^N \mathbb{E}\big|\Delta_s L(Y) - \mathbb{E}\Delta_s L(Y)\big|.$$

Before applying the result, we make a couple of general remarks to aid in interpretation.

REMARK 1.2.   Regarding the conditions of the theorem: writing the probabilities of X in the form (1.1) is without loss of generality assuming $\mathbb{P}(X = x) > 0$ for all $x \in \{0, 1\}^N$ (or we allow L to take the value $-\infty$). The condition (1.2) comes from matching the update probabilities of the Glauber dynamics of $\mathscr{L}(X)$ to those of the independent chain, while respecting the behaviour of the statistic in the exponent under the product measure; this is the so-called mean field prediction (cf. Eldan and Gross (2018a)). Solutions to this equation also arise as critical points for variational problems arising from studying large deviations in these systems; see Ellis (2006), Chatterjee and Diaconis (2013), and discussion below.

Chains with a coupling satisfying (1.3) are called *contracting* in Gheissari, Lubetzky and Peres (2018), and the condition is satisfied by many chains in statistical physics models in "high temperature" regimes; see Levin, Peres and Wilmer ((2009), Chapters 14 and 15) and applications below. The condition is implied by the Dobrushin–Shlosman condition, going back to Dobrushin and Shlosman (1985), which is a classical criterion for fast mixing of the Glauber dynamics, among other nice implications. Our general result, Theorem 2.1, replaces the Dobrushin–Shlosman condition (1.3) by the more general assumption that the "influence matrix" of the Glauber dynamics has p-norm less than 1 (Dobrushin–Shlosman corresponds to $p = 1$). Such an assumption still implies fast mixing of the Glauber dynamics; see Dobrushin (1970), Hayes (2006) and Dyer, Goldberg and Jerrum (2009).

Also note that under the assumptions of the theorem, the bound is easy to compute since the expectation in (1.4) is taken against a vector Y of *independent* variables. A key to our approach is that the intractable normalising constant κ of $\mathscr{L}(X)$ does not need to be computed in order to analyse the Glauber dynamics.

REMARK 1.3.   In a nice situation, $\rho \asymp 1/N$, $\mathbb{E}|\Delta_s L(Y) - \mathbb{E}\Delta_s L(Y)| = O(1/\sqrt{N})$, and the bound is then $\|\Delta h\| O(\sqrt{N})$. The way to interpret the bound is that for functions h such that $\|\Delta h\|\sqrt{N} \ll \mathbb{E}h(Y)$, $\mathbb{E}h(X)$ can be approximated by $\mathbb{E}h(Y)$ with small error. If $\mathbb{E}h(Y)$ is of constant order with N, then the function h must satisfy $\|\Delta h\| = o(1/\sqrt{N})$ for the bound to be meaningful.

REMARK 1.4.   In the case that the chain does not satisfy the mixing assumption (1.3), or that of the more general Theorem 2.1, Lemmas 2.4 and 2.5 give an intermediate bound that can be used directly, as is done to prove Theorem 1.13 below.

We apply Theorem 1.1 in some examples and applications.

1.1. *Ising model on a fixed graph.* Let G be a fixed graph on N vertices with vertex set $V(G) = [N]$ and edge set $E(G)$. Each vertex has a label in $\{0, 1\}$ attached to it, and a configuration of such labels is denoted by $x \in \{0, 1\}^N$. Let $r \in [N - 1]$ and for each $s \in [N]$, let $\mathcal{N}_s \subset [N] \setminus \{s\}$ be a nonrandom set of size $|\mathcal{N}_s| \le r$, such that $u \in \mathcal{N}_s \iff s \in \mathcal{N}_u$. One can think of r as the largest vertex degree in G. Let $X \in \{0, 1\}^N$ have the Ising model with "neighbourhoods" $(\mathcal{N}_s)_{s \in [N]}$, defined by $\mathbb{P}(X = x) \propto e^{L(x)}$, where

$$L(x) = \frac{\beta}{N} \sum_{s=1}^{N} \sum_{t \in \mathcal{N}_s} (2x_s - 1)(2x_t - 1).$$

To apply Theorem 1.1, we first compute

$$\Delta_s L(x) = \frac{\beta}{N} \left\{ \sum_{t \in \mathcal{N}_s} (2 - 1)(2x_t - 1) - \sum_{t \in \mathcal{N}_s} (0 - 1)(2x_t - 1) \right.$$

$$\left. + \sum_{u : u \neq s} \mathbb{1}(s \in \mathcal{N}_u) \{ (2x_u - 1)(2 - 1) - (2x_u - 1)(0 - 1) \} \right\}$$

$$= 4 \frac{\beta}{N} \sum_{t \in \mathcal{N}_s} (2x_t - 1),$$

where we used that $u \in \mathcal{N}_s \iff s \in \mathcal{N}_u$. Hence

$$\Delta_s L(x) = \frac{4\beta}{N} \sum_{t \in \mathcal{N}_s} (2x_t - 1).$$

For $(Y_s)_{s \in [N]}$, a vector of independent Bernoulli variables with $\mathbb{P}(Y_s = 1) = p_s$,

$$\mathbb{E} \Delta_s L(Y) = \frac{4\beta}{N} \sum_{t \in \mathcal{N}_s} (2p_t - 1),$$

and according to (1.2), we set the p_s to satisfy

(1.5) $$(2p_s - 1) = \tanh\left(\frac{2\beta}{N} \sum_{t \in \mathcal{N}_s} (2p_t - 1) \right).$$

Equation (1.5) always has at least one solution: $p_s \equiv a = 1/2$. For the mixing time, we restrict to the high temperature regime where $0 < \beta < r/N$. According to Levin, Peres and Wilmer ((2009), Proof of Theorem 15.1), with their β corresponding to our β/N (and see also Levin, Luczak and Peres ((2010), Proposition 2.1) for the case $r = (N - 1)$), there is a coupling of $(U^{[x,s]}(1), V^{[x,s]}(1))$ such that

$$\mathbb{E} d_H(U^{[x,s]}(1), V^{[x,s]}(1)) \le (1 - N^{-1}(1 - r \tanh(\beta/N))),$$

so that in applying Theorem 1.1, we can set $\rho = N^{-1}(1 - r\tanh(\beta/N)) \geq (1 - \beta r/N)N^{-1}$; since tanh is 1-Lipschitz. Hence Theorem 1.1 implies that or any $h : \{0,1\}^N \to \mathbb{R}$,

$$(1.6) \qquad |\mathbb{E}h(X) - \mathbb{E}h(Y)| \leq \frac{\|\Delta h\|}{4(1 - \beta r/N)} \sum_{s=1}^{N} \mathbb{E}|\Delta_s L(Y) - \mathbb{E}\Delta_s L(Y)|.$$

Assuming $p_s \equiv a$, it is easy to see that $\Delta_s L(Y) \overset{d}{=} 8\beta N^{-1} B_s + c$, where $B_s \sim \text{Bi}(|\mathcal{N}_s|, a)$ and $c = -4\beta|\mathcal{N}_s|/N$ is a constant. So with $a = \frac{1}{2}$ and Y a vector of i.i.d. Bernoulli(1/2) random variables,

$$\mathbb{E}|\Delta_s L(Y) - \mathbb{E}\Delta_s L(Y)| \leq \sqrt{\text{Var}(\Delta_s L(Y))} = \frac{8\beta\sqrt{|\mathcal{N}_s|a(1-a)}}{N} \leq \frac{4\beta\sqrt{r}}{N}.$$

Combining this and (1.6), Theorem 1.1 implies that for any $h : \{0,1\}^N \to \mathbb{R}$ and $0 \leq \beta < r/N$,

$$(1.7) \qquad |\mathbb{E}h(X) - \mathbb{E}h(Y)| \leq \|\Delta h\| \frac{\beta\sqrt{r}}{(1 - \beta r/N)}.$$

To put this result in context, the model with $|\mathcal{N}_s| = (N - 1)$ is frequently referred to as the Curie–Weiss model and is one of the most well studied in statistical physics. Laws of large numbers, central limit theorems with rates, large deviations, concentration and moment inequalities and local limit theorems with rates have been established for this model; see Ellis and Newman (1978), Ellis, Newman and Rosen (1980), Ellis (2006), Chatterjee (2007), Eichelsbacher and Löwe (2010), Chatterjee and Shao (2011), Röllin and Ross (2015), Barbour, Röllin and Ross (2019). However, it is not obvious how to get an explicit result like (1.7) directly from these results. Some very recent work closely related to (1.7) is Eldan and Gross ((2018a), Corollary 3.5), which implies that for $\beta < 1$,

$$(1.8) \qquad |\mathbb{E}h(X) - \mathbb{E}h(Y)| \leq \|\Delta h\| \left(\frac{5001(1 + \beta)^2}{2(1 - \beta)} + 12 \right) N^{7/8}.$$

Our bound (1.7) in this case is essentially $\|\Delta h\|\beta\sqrt{N}/(1 - \beta)$, which compares favorably to (1.8) in rate, constant and dependence on β (i.e., goes to zero with β).

1.2. *Exponential random graph models.* Exponential random graph models (ERGMs), suggested for directed networks by Holland and Leinhardt (1981) and for undirected networks by Frank and Strauss (1986), are frequently used as parametric statistical models in network analysis; see, for example, Wasserman and Faust (1994). This is due to the Gibbs form of the distribution, which allows for straightforward implementation of modern Markov chain Monte Carlo (MCMC) methods. However, the models are difficult to analyse directly, and so the stability of MCMC algorithms and the structure of the resulting networks have only recently been meaningfully studied. Two important references, both in general and

for this work, are Bhamidi, Bresler and Sly (2011), which establishes mixing times for the MCMC (Gibbs samplers) used for ERGMs, and Chatterjee and Diaconis (2013), which establishes asymptotic properties of these random networks. A further discussion is postponed to after the statement of our definitions and results.

Of particular interest in the analysis of networks are counts of small graphs. These counts occur in the theory of graph limits; see Section 2.2 in Chatterjee and Diaconis (2013) for an overview. Over-and under-represented small subgraphs are conjectured to be building blocks of complex networks; see, for example, Milo et al. (2002). Small subgraphs are used for comparing networks; see Pržulj (2007), Ali et al. (2014). Assessing the exceptionality of small subgraph counts, also called *motifs*, depends crucially on the underlying network model; see Picard et al. (2008). Hence assessing the distribution of small subgraph counts in ERGMs is a key question. While our results do not provide the distribution of these counts, they provide an approximation in terms of the well-studied corresponding distributions for Bernoulli random graphs, together with a bound on the error in the approximation which allows to gauge whether the approximation is appropriate.

We first define the vertex-labeled ERGM. Let $\mathcal{G}_n^{\text{lab}}$ be the set of vertex-labeled simple graphs on n vertices and define the set $[n]_2 := \{(i, j) : 1 \leq i < j \leq n\}$ (this is sometimes also denoted $\binom{[n]}{2}$ but we prefer our notation for typesetting purposes). We identify $\mathcal{G}_n^{\text{lab}}$ with $\{0, 1\}^{\binom{n}{2}}$ by encoding $x \in \mathcal{G}_n^{\text{lab}}$ by an ordered collection of $0 - 1$ valued variables: $x = (x_{ij})_{(i,j)\in[n]_2}$, where $x_{ij} = 1$ means there is an edge between vertices i and j. We refer to x_{ij} as the ijth "coordinate" of x. Note that in the general setup above, $N = \binom{n}{2}$.

For a graph H, let $V(H)$ denote the vertex set, and for $x \in \{0, 1\}^{\binom{n}{2}}$, define $t(H, x)$ to be the number of "edge-preserving" injections from $V(H)$ to $V(x)$; an injection σ preserves edges if for all edges vw of H, $x_{\sigma(v)\sigma(w)} = 1$ (here assuming $\sigma(v) < \sigma(w)$).

DEFINITION 1.5 (Exponential random graph model). Fix $n \in \mathbb{N}$ and k connected graphs $H_1, \ldots, H_k$ with H_1 a single edge, and for $\ell = 1, \ldots, k$ denote $v_\ell := |V(H_\ell)|$ (so $v_1 = 2$), and

$$t_\ell(x) = \frac{t(H_\ell, x)}{n(n-1)\cdots(n-v_\ell+3)}.$$

For $\beta = (\beta_1, \ldots, \beta_k)$ with $\beta_\ell \in \mathbb{R}$ for $\ell = 1, \ldots, k$, we say the random graph $X \in \mathcal{G}_n^{\text{lab}}$ is distributed according to the exponential random graph model with parameters β, denoted $X \sim \text{ERGM}(\beta)$, if for $x \in \mathcal{G}_n^{\text{lab}}$,

$$\mathbb{P}(X = x) = \frac{1}{\kappa_n(\beta)} \exp\left(\sum_{\ell=1}^{k} \beta_\ell t_\ell(x)\right),$$

where $\kappa_n(\beta)$ is a normalizing constant.

The scaling in the exponent matches Bhamidi, Bresler and Sly ((2011), Definition 1) and Chatterjee and Diaconis ((2013), Sections 3 and 4). Note also that $t_1(H_1, x)$ is twice the number of edges of x so that if $k = 1$, then ERGM(β) has the same law as an Erdős–Rényi random graph (identified above as a collection of $\binom{n}{2}$ i.i.d. Bernoulli variables) with edge parameter $e^{2\beta_1}/(1 + e^{2\beta_1})$.

As in Bhamidi, Bresler and Sly (2011) and Eldan and Gross (2018b), an important role is played by the following functions defined on $[0, 1]$,

$$\Phi(a) := \sum_{\ell=1}^{k} \beta_\ell e_\ell a^{e_\ell - 1} \quad \text{and} \quad \varphi(a) := \frac{1 + \tanh(\Phi(a))}{2} = \frac{e^{2\Phi(a)}}{e^{2\Phi(a)} + 1},$$

where e_ℓ is the number of edges in H_ℓ. In particular, solutions to the equation $\varphi(a) = a$ satisfying $\varphi'(a) < 1$ are key quantities in what follows. To help understand where these equations come from, Chatterjee and Diaconis ((2013), Theorem 4.2), states that if $\beta_\ell \geq 0$ for $\ell = 2, \ldots, k$, then $X \sim$ ERGM(β) is asymptotically close (in the cut metric) to a mixture of Erdős–Rényi random graphs where the mixture is over the finite set U of maximizers in the interval $[0, 1]$, of the function

$$(1.9) \qquad \sum_{\ell=1}^{k} \beta_\ell a^{e_\ell} - \frac{1}{2}\big(a \log(a) + (1 - a) \log(1 - a)\big).$$

Moreover, Chatterjee and Diaconis (2013) show that the set U is finite, and possibly only has one element. The connection to solutions of $\varphi(a) = a$ is that critical points of (1.9) satisfy

$$(1.10) \qquad 2\Phi(a) = \log\left(\frac{a}{1 - a}\right),$$

which a little algebra shows is the same as $\varphi(a) = a$. The second condition that $\varphi'(a) < 1$ corresponds to the critical point being a local maximum. Thus, such solutions are key to describing any Erdős–Rényi limiting behaviour of ERGMs.

It is intuitively clear that if $\beta_2, \ldots, \beta_k$ are small, then ERGM(β) should be close to an Erdős–Rényi random graph. Our first ERGM result explicitly quantifies this heuristic. Here and below, we denote

$$C_2 := \frac{4}{3\sqrt{3}} < 0.77,$$

which appears in our study as the maximum of the first derivative of $\text{sech}^2(a)$. Moreover, from the next theorem onwards, Z always has the Erdős–Rényi distribution with parameter a^*, where $a^* \in [0, 1]$ satisfies $a^* = \varphi(a^*)$.

THEOREM 1.6. *For given $\beta_1 \in \mathbb{R}$, $\beta_\ell > 0$, $\ell = 2, \ldots, k$, assume $a^* \in [0, 1]$ satisfies $a^* = \varphi(a^*)$, define $A^* := \max\{a^*, 1 - a^*\} \leq 1$, set*

$$\alpha_1 := \frac{1}{2}\big(\Phi'(a^*) + A^* \Phi''(1)\big),$$

$$\alpha_2 := \varphi'(a^*)$$
$$+ \frac{1}{2}[C_2(A^* + n^{-1})\Phi'(1)(\Phi'(a^*) + A^*\Phi'(1)) + A^*\Phi''(1)\operatorname{sech}^2(\Phi(a^*))],$$

and assume

(1.11) $$1 - \min\{\alpha_1, \alpha_2\} =: \gamma > 0.$$

Then for $X \sim \mathrm{ERGM}(\beta)$, Z having the Erdős–Rényi distribution with parameter a^*, and $h : \{0, 1\}^{\binom{n}{2}} \to \mathbb{R}$,

$$\left|\mathbb{E}h(X) - \mathbb{E}h(Z)\right| \le \|\Delta h\| \binom{n}{2}(4\gamma)^{-1}\sum_{\ell=2}^{k}\beta_\ell\sqrt{\mathrm{Var}(\Delta_{12}t_\ell(Z))}.$$

The condition (1.11) may look contrived, but note the theoretically nice fact that a necessary condition for $1 - \alpha_2 > 0$, is that $\varphi'(a^*) < 1$, and this corresponds to the critical point a^* of (1.9) being a local maximum, as explained above. We now prove a similar result, which has simpler conditions and allows β_ℓ to be negative for $\ell \ge 2$. Showing how our methods can be used to extend to this case is due to Sinulis ((2018), Proofs of Theorem 2.2 and Proposition 2.4), which appeared after the first draft of this paper was put on the arXiv. Define $|\Phi|(a) := \sum_{\ell=1}^{k}|\beta_\ell|e_\ell a^{e_\ell - 1}$.

THEOREM 1.7. *For given $\beta_\ell \in \mathbb{R}$, $\ell = 1, \ldots, k$, assume that $\frac{1}{2}|\Phi|'(1) < 1$ and that $a^* \in [0, 1]$ satisfies $a^* = \varphi(a^*)$. Then for $X \sim \mathrm{ERGM}(\beta)$, Z having the Erdős–Rényi distribution with parameter a^*, and $h : \{0, 1\}^{\binom{n}{2}} \to \mathbb{R}$,*

$$\left|\mathbb{E}h(X) - \mathbb{E}h(Z)\right| \le \|\Delta h\| \binom{n}{2}\left(4\left(1 - \frac{1}{2}|\Phi|'(1)\right)\right)^{-1}\sum_{\ell=2}^{k}|\beta_\ell|\sqrt{\mathrm{Var}(\Delta_{12}t_\ell(Z))}.$$

REMARK 1.8. The condition from Theorem 1.7,

(1.12) $$|\Phi|'(1) < 2,$$

is easier to verify than (1.11) of Theorem 1.6. Moreover, the dependence on β of (1.12) is more transparent than (1.11), since it does not involve the fixed point a^*. In particular, it is plain to see that for any fixed choice of $H_2, \ldots, H_k$, (1.12) holds for $|\beta|$ small enough.

REMARK 1.9. For fixed a^*, the random variables

$$\Delta_{(1,2)}t_\ell(Z) = \frac{\Delta_{(1,2)}t(H_\ell, Z)}{n(n-1)\cdots(n - v_\ell + 3)},$$

have variance of order at most $1/n$, with constant depending on properties of H_ℓ.

To see this, let H and x be graphs with vertex and edge sets $V(H)$, $V(x)$, $E(H)$, $E(x)$, and let $I(H, x)$ be the set of all injections $i_{H,x} : V(H) \to V(x)$. For such an injection and any edge $e = (u, v)$ of H, we use the notation $i_{H,x}(e) = \{i_{H,x}(u), i_{H,x}(v)\}$ and $E(i_{H,x}) = \bigcup_{e \in E(H)}\{i_{H,x}(e)\}$. Then the number of edge-preserving injections of $I(H, x)$ is

$$t(H, x) = \sum_{i_{H,x} \in I(H,x)} \prod_{e \in E(H)} \mathbb{1}\big(i_{H,x}(e) \in E(x)\big),$$

and so for $s \in E(x)$,

$$\Delta_s t(H, x)$$

$$= \sum_{i_{H,x} \in I(H,x)} \left\{ \prod_{e \in E(H)} \mathbb{1}\big(i_{H,x}(e) \in E(x^{(s,1)})\big) - \prod_{e \in H} \mathbb{1}\big(i_{H,x}(e) \in E(x^{(s,0)})\big) \right\}$$

$$= \sum_{i_{H,x} \in I(H,x)} \mathbb{1}\big(s \in E(i_{H,x})\big) \prod_{e \in E(H) \setminus i_{H,x}^{-1}(s)} \mathbb{1}\big(i_{H,x}(e) \in (E(x) \setminus i_{H,x}(s))\big).$$

Now, $|I(H, x)|$ is of the order $n^{|v(H)|}$ and the number of such edge-preserving injections $i_{H,x}$ which use the edge s, so that $s \in E(i_{H,x})$, is of the order $n^{|v(H)|-2}$. The variance of $\Delta_s t(H, Z)$ involves of the order $n^{2(|v(H)|-2)}$ covariances, but many of these covariances will be 0 because edges in Z are independent. Indeed, a nonzero covariance is obtained only when the two injections share at least one edge, which leaves only at most $|v(H)| - 3$ vertices to choose from for the injection. Hence

$$\text{Var}\,\Delta_s t(H, Z) = O\big(n^{2|v(H)|-5}\big).$$

Thus taking the denominator in $\Delta_{(1,2)} t_\ell(Z)$ into account yields $\text{Var}\,\Delta_{(1,2)} t_\ell(Z) = O(n^{-1})$. This argument has been explored for subgraph counts in Bernoulli graphs in many previous works, for example, Ruciński (1988).

Special test functions of interest are homomorphism densities $h(x) = t(H, x) \times n^{-|V(H)|}$ for small nontrivial subgraphs H, since convergence of subgraph densities plays a key role in many other notions of convergence; see Borgs et al. (2008, 2012), and as mentioned previously, counts of small subgraphs are used in the analysis of networks as building blocks and summary statistics. We have the following corollary, which follows easily from Theorems 1.6 and 1.7 and Remark 1.9 after noting that $\|\Delta h\| = O(n^{-2})$, providing a rate of convergence for this case.

COROLLARY 1.10. *If β satisfies the hypotheses of Theorems 1.6 or 1.7, and for some graph H, $h(x) = t(H, x) n^{-|V(H)|}$, then for $X \sim \text{ERGM}(\beta)$, Z having*

the Erdős–Rényi distribution with parameter a^*,

$$\left|\mathbb{E}h(X) - \mathbb{E}h(Z)\right| \leq \frac{c}{\sqrt{n}},$$

where c is a computable constant depending on β and H.

Explicit computation of the region of applicability of Theorem 1.7 is possible for small—though frequently used (see Wasserman and Faust (1994))—examples, as we now illustrate.

PROPOSITION 1.11. *Let $X \sim \mathrm{ERGM}(\beta)$ with $k = 2$, H_2 a two-star, $\beta_1 \in \mathbb{R}$, and $|\beta_2| < 1$. Then there is a unique $a^* \in [0, 1]$ satisfying $a^* = \varphi(a^*)$, and for Z distributed Erdős–Rényi with parameter a^*, and any $h : \{0, 1\}^{\binom{n}{2}} \to \mathbb{R}$,*

$$\left|\mathbb{E}h(X) - \mathbb{E}h(Z)\right| \leq \|\Delta h\| \binom{n}{2} \left(4\left(1 - |\beta_2|\right)\right)^{-1} \frac{\sqrt{8a^*(1 - a^*)}|\beta_2|}{\sqrt{n - 2}}.$$

PROPOSITION 1.12. *Let $X \sim \mathrm{ERGM}(\beta)$ with $k = 2$, H_2 be a triangle, $\beta_1 \in \mathbb{R}$, and $|\beta_2| < 1/3$. Then there is a unique $a^* \in [0, 1]$ satisfying $a^* = \varphi(a^*)$, and for Z distributed Erdős–Rényi with parameter a^* and any $h : \{0, 1\}^{\binom{n}{2}} \to \mathbb{R}$,*

$$\left|\mathbb{E}h(X) - \mathbb{E}h(Z)\right| \leq \|\Delta h\| \binom{n}{2} \left(4\left(1 - 3|\beta_2|\right)\right)^{-1} \frac{6a^*\sqrt{1 - (a^*)^2}|\beta_2|}{\sqrt{n - 2}}.$$

For comparison, if $k = 2$ and H_2 is a triangle, the recent result of Eldan and Gross ((2018b), Theorem 19) states that in the range $0 \leq |\beta_2| < 1/3$,

$$\left|\mathbb{E}h(X) - \mathbb{E}h(Z)\right| \leq \|\Delta h\| \frac{5000((72|\beta_2|) \vee 2)^2}{1 - 3|\beta_2|} \binom{n}{2}^{15/16}.$$

Our bound compares favourably to this one in rate and constant.

Before stating our next ERGM result and having a further discussion, we give a quick toy numerical example.

Florentine marriages. The Florentine Families Marriage data is a classical data set collected by Padgett and Ansell (1993). It is an undirected network, which consists of the marriage ties among 16 families in 15th century Florence, Italy; each family is a vertex, and two vertices are connected by an edge if there is a marriage between them. There are $n = 16$ vertices and 20 edges in the network, and one of the vertices is isolated; the network has $50 * 2 = 100$ subgraphs which are isomorphic to a 2-star. Figure 1 shows this network; the vertices are labelled by the family names.
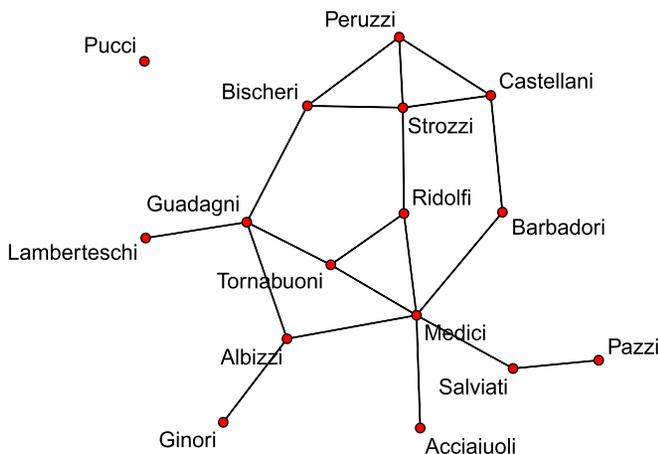
FIG. 1. *Marriage relations between Florentine families.*

Using the R package *sna*,[3] we fit an exponential random graph model to these data with $k = 2$ and H_2 a 2-star. The fitted coefficients are $\beta_1 = -1.6339$ and $\beta_2 = 0.0049 * 2 = 0.0098$, and so Proposition 1.11 applies. The function Φ is

$$\Phi(a) = -1.6339 + 0.0196a,$$

and the unique solution to $\varphi(a) = a$ is $a^* = 0.036743$. Thus the upper bound of Proposition 1.11 between the fitted model and an Erdős–Rényi graph with parameter a^* is approximately

$$\binom{16}{2} \frac{0.036743 \sqrt{8(0.0098)(1 - 0.036743)}}{(4(1 - 0.0098))\sqrt{14}} = 0.0817595.$$

Our next result is a generalisation of Theorem 1.6, which applies to the whole high temperature regime, meaning there is a unique critical point for (1.9) and it is a maximum. The result is more general, but the constants are not explicit.

THEOREM 1.13. *For given $\beta_1 \in \mathbb{R}$, $\beta_\ell > 0$, $\ell = 2, \ldots, k$, assume there is a unique $a^* \in [0, 1]$ that satisfies $a^* = \varphi(a^*)$, and that this unique fixed point satisfies $\varphi'(a^*) < 1$. Then for $X \sim \mathrm{ERGM}(\beta)$, Z having the Erdős–Rényi distribution with parameter a^*, and any function $h : \{0, 1\}^{\binom{n}{2}} \to \mathbb{R}$, there is a constant C depending only on β and $H_1, \ldots, H_\ell$, such that*

$$\big| \mathbb{E}h(X) - \mathbb{E}h(Z) \big| \le C \|\Delta h\| n^{3/2}.$$

[3]https://cran.r-project.org/web/packages/sna/sna.pdf

REMARK 1.14.   To interpret this bound, note that setting $\|\Delta h\| = 1$ gives an upper bound on the Wasserstein distance (with respect to Hamming distance) between $\mathscr{L}(X)$ and $\mathscr{L}(Z)$. An alternative definition of the Wasserstein distance is

$$\inf_{\substack{(\hat{X}, \hat{Z}) \text{ s.t.} \\ \hat{X} \overset{d}{=} X, \hat{Z} \overset{d}{=} Z}} \mathbb{E} \, d_{\mathrm{H}}(\hat{X}, \hat{Z}),$$

and so the bound implies we can couple X to Z in such a way that on average there are $\mathrm{O}(n^{3/2})$ edges that do not match. The bound implies a meaningful approximation since there are $\mathrm{O}(n^2)$ edges in X.

Now, to compare our ERGM results to those existing, we already mentioned the result of Chatterjee and Diaconis (2013), which shows large ERGMs are close to finite mixtures of Erdős–Rényi random graphs. Theorems 1.6 and 1.7 explicitly quantify this in a subset of the high temperature regime, while Theorem 1.13 gives rates over the whole high temperature regime. Our results rely on a close analysis of the ERGM Glauber dynamics, following Bhamidi, Bresler and Sly (2011), where mixing times are studied. In particular, they find that in the high temperature regime, the chain mixes at the same order ($n^2 \log(n)$) as the Erdős–Rényi case. The region described in our Theorem 1.6 is a subset of the high temperature regime where the Glauber dynamics are contracting. For parameter values where there is more than one *local* maximum, Bhamidi, Bresler and Sly (2011) show that the chain takes exponential (in n) time to mix. One approach to adapt our results to this case is to study the Glauber dynamics of the ERGM conditioned to be in a region "close" to a local maximum where the dynamics are contracting; a suitable region can be read from Lemma 3.1. We also mention Radin and Yin (2013), where the variational problem (1.9) is analysed in the case when $k = 2$ and $\beta_2 > 0$ or where $\beta_2 < 0$ in the case where H_2 is a star; even this basic case is important to understand, though not straightforward to analyse. Further analysis of the variational problem for $k = 3$ is given in Yin (2013).

Closest to our ERGM results is Eldan and Gross (2018b) (already mentioned in an example above), and in particular they showed in their Theorem 18 that for positive β_ℓ for $\ell = 2, \ldots, k$ and assuming there is a unique solution a^* to (1.10) and with $\varphi'(a^*) < 1$ (the high temperature regime) then for $X \sim \mathrm{ERGM}(\beta)$, Z having the Erdős–Rényi distribution with parameter a^*, and any $h : \{0, 1\}^{\binom{n}{2}} \to \mathbb{R}$,

$$\left| \mathbb{E} h(X) - \mathbb{E} h(Z) \right| \leq \|\Delta h\| \mathrm{O}(n^{2(1-\theta)})$$

for some $0 < \theta < 1/16$, and the constant can be made numerically explicit in fixed examples. The result is more general than Theorem 1.6, as it covers the entire high temperature regime, but the rate is not as good. Theorem 1.13 covers the high temperature regime with a better rate than Eldan and Gross ((2018b), Theorem 18), but without explicit constants (and really no hope of extracting them from our proof).

The remainder of this paper is organized as follows. We finish the Introduction with a summary discussion. In Section 2, we state and prove our main general approximation result, Theorem 2.1, and show how it implies Theorem 1.1. In Section 3, we state and prove our ERGM results, Theorems 1.6 and 1.7, and Propositions 1.11 and 1.12 and Theorem 1.13.

1.3. *Discussion.* Our main result gives an easy to compute upper bound between a vector of Bernoulli variables that are the stationary distribution of a fast mixing Markov chain and an independent vector. In typical applications, such as Ising models and ERGMs, the result gives approximation with rates of convergence for expectations of certain statistics of the system. Our work further supports the notion, appearing in other contexts, that stationary distributions of fast mixing Markov chains should have properties similar to independent systems. Our bounds are simple and explicit, and improve upon others given in Eldan and Gross (2018a, 2018b), which in turn are derived from a "nonlinear" large deviations framework developed in Eldan (2018). That framework uses a notion of complexity referred to as *Gaussian width*. A similar framework with a different notion of complexity is given in Chatterjee and Dembo (2016); see also Chatterjee (2016). It would be interesting to make the connection between these complexity measures and fast mixing explicit, which would complete more of the picture around concentration, mixing, and distance to independence.

Though our results only apply in high temperature regimes, note that in low temperature regimes, fast mixing may occur for the measure conditioned to be in a "good" neighbourhood. For example, in the ERGM, we would condition on being in the set given by (3.2) for small enough ε such that (3.3) is satisfied. The difficulty is to ensure that the conditional Glauber dynamics are well behaved at the boundary.

REMARK 1.15. After the second author spoke about this work at the MIT probability seminar on 23 October 2017, we became aware that Guy Bresler and Dheeraj Nagaraj have been working independently on a similar general approximation result, with applications to Ising models; see Bresler and Nagaraj (2017). In particular, they prove Lemmas 2.3 and 2.4 using roughly the same ideas as ours, and explore their use in Ising models, going beyond the illustrative Section 1.1. Results for ERGMs, which are our main application, are not developed in their work. For added transparency, we have coordinated the submissions of our two papers.

2. General approximation result. In this section, we state and prove our general approximation result, and then show how it implies Theorem 1.1. As previously mentioned, we work in the setup of Hayes (2006) and Dyer, Goldberg and

Jerrum (2009). Define the $N \times N$ *influence matrix* $\hat{R}$ for the Glauber dynamics of $\mathscr{L}(X)$ by

$$\hat{R}_{rs} := \max_{x \in \{0,1\}^N} |q_X((x^{(s,1)})^{(r,1)}|x^{(s,1)}) - q_X((x^{(s,0)})^{(r,1)}|x^{(s,0)})|.$$

Then $\hat{R}_{rs}$ is the maximum amount that the conditional distribution of the rth coordinate of x can change due to a change in the sth coordinate of x. For $1 \leq p \leq \infty$, let $\| \cdot \|_p$ be the p-norm on $\mathbb{R}^N$, and define the matrix operator p-norm

$$\|A\|_p := \sup_{v \neq 0} \frac{\|Av\|_p}{\|v\|_p};$$

as is typical, the notation $\| \cdot \|_p$ has different meaning depending on whether the argument is a matrix or vector, but this should not cause confusion.

THEOREM 2.1. *Let $X, Y \in \{0,1\}^N$ be random vectors, $h : \{0,1\}^N \to \mathbb{R}$, and assume the continuous time Glauber dynamics for $\mathscr{L}(X)$ is irreducible. For $s \in [N]$, set $c_s := \|\Delta_s h\|$ and $v_s(y) := |q_X(y^{(s,1)}|y) - q_Y(y^{(s,1)}|y)|$, and $c := (c_1, \ldots, c_N)$ and $v(Y) := (v_1(Y), \ldots, v_N(Y))$. Assume there is an $N \times N$ matrix R, such that for all $r, s \in [N]$, and some $1 \leq p \leq \infty$ and $\varepsilon = \varepsilon_p > 0$,*

(2.1) $$\hat{R}_{rs} \leq R_{rs} \quad and \quad \|R\|_p \leq 1 - \varepsilon < 1,$$

where $\hat{R}$ is the influence matrix for the Glauber dynamics of $\mathscr{L}(X)$. Then for $q := p/(p-1)$,

$$|\mathbb{E}h(X) - \mathbb{E}h(Y)| \leq \varepsilon^{-1} \|c\|_q \mathbb{E}\|v(Y)\|_p.$$

REMARK 2.2. A matrix R such that (2.1) holds is also called a *dependency matrix*. The condition $\|R\|_p \leq 1 - \varepsilon < 1$ with $p = \infty$ (maximum of row sums) is the *Dobrushin* condition from Dobrushin (1970), and with $p = 1$ (maximum of column sums) is the *Dobrushin–Shlosman* condition from Dobrushin and Shlosman (1985). Either condition implies fast mixing of Glauber dynamics, and it was shown in Hayes (2006) that (2.1) with $p = 2$ implies fast mixing of Glauber dynamics (even with "random update"). Moreover, from Dyer, Goldberg and Jerrum (2009), the same bound for more general matrix norms also implies rapid mixing of Glauber dynamics (it may be possible to adapt the methods there to obtain a variation of the bound of Theorem 2.1). These papers cover a variety of applications, for example, Ising and hard core models and proper graph colorings, where Theorem 2.1 can be applied. We pursue further applications, generalisations and their implications elsewhere.

The proof of the theorem uses Stein's method, due to Stein (1972, 1986); see Barbour and Chen (2005), Chen, Goldstein and Shao (2011), Ley, Reinert and

Swan (2017), Ross (2011) for various introductions; and we follow Barbour's generator method from Barbour (1988, 1990). Our approach is related to ideas from Reinert (2005) for functions of independent variables, and Eichelsbacher and Reinert (2008) for one-dimensional Gibbs measures. The generator approach requires bounds on certain mixing quantities, and these are obtained from the bound on the norm of the influence matrix, following Hayes (2006), Dyer, Goldberg and Jerrum (2009) and Bubley and Dyer (1997). We also remark that the key Lemma 2.5 below is closely related to ideas in Chatterjee ((2005), Section 4.2), which gives fast mixing criteria for concentration of Lipschitz functions of stationary distributions.

The underlying idea to the generator approach is that a distribution μ is characterised as the stationary distribution of a Markov process which has generator $\mathcal{A}_\mu$. The generator $\mathcal{A}$ appears in the so-called *Stein equation*

$$\mathcal{A}_\mu f(x) = h(x) - \mathbb{E}h(X),$$

where X has distribution μ and h is a test function. Two distributions can then be compared through the comparison of the corresponding generators; if Y has distribution ν then

$$\mathbb{E}h(Y) - \mathbb{E}h(X) = \mathbb{E}\mathcal{A}_\mu f(Y) = \mathbb{E}\mathcal{A}_\mu f(Y) - \mathbb{E}\mathcal{A}_\nu f(Y).$$

Here, the last equality follows from $\mathbb{E}\mathcal{A}_\nu f(Y) = 0$ under some regularity assumptions.

A Markov process which has stationary distribution μ can sometimes be obtained through a sequence of exchangeable operations on the configurations which the chain can take on; see, for example, Rinott and Rotar (1997) and Reinert (2005). In this paper, the exchangeable construction is provided by the Glauber dynamics with an exponential clock for the transition times.

The first lemma we need in order to prove the theorem is by now standard in Stein's method; it provides a solution of the Stein equation induced by the generator of the Glauber dynamics. Define the continuous time Glauber dynamics with exponential rate 1 holding times, having generator $\mathcal{A} := \mathcal{A}_X$ given by

$$(2.2) \qquad \mathcal{A}f(x) = \frac{1}{N} \sum_{s \in [N]} \left[q(x^{(s,1)}|x)\Delta_s f(x) + (f(x^{(s,0)}) - f(x)) \right].$$

LEMMA 2.3. *Let $X \in \{0, 1\}^N$ be a random vector and $(X(t))_{t \geq 0}$ be the continuous time Glauber dynamics Markov chain for $\mathscr{L}(X)$ with generator (2.2). If the Markov chain is irreducible, then for any $h : \{0, 1\}^N \to \mathbb{R}$, the function*

$$f_h(x) := -\int_0^\infty \mathbb{E}\left[h(X(t)) - \mathbb{E}h(X)|X(0) = x\right] dt$$

is well defined and satisfies $\mathcal{A}f_h(x) = h(x) - \mathbb{E}h(X)$.

The next lemma gives an intermediate bound on the quantity we want to approximate, through comparing the generators of the Glauber dynamics (2.2) for the two processes under investigation.

LEMMA 2.4. *Let* $X, Y \in \{0, 1\}^N$ *be random vectors,* $h : \{0, 1\}^N \to \mathbb{R}$, f_h *be as in Lemma* 2.3. *Then*

$$\big|\mathbb{E}h(X) - \mathbb{E}h(Y)\big| \leq \frac{1}{N} \sum_{s \in [N]} \mathbb{E}\big[\big|q_X(Y^{(s,1)}|Y) - q_Y(Y^{(s,1)}|Y)\big|\big|\Delta_s f_h(Y)\big|\big].$$

PROOF. From Lemma 2.3 and the fact that $\mathbb{E}\mathcal{A}_Y f(Y) = 0$, we have

$$\begin{aligned}
\big|\mathbb{E}h(X) - \mathbb{E}h(Y)\big| &= \big|\mathbb{E}\mathcal{A}_X f_h(Y)\big| \\
&= \big|\mathbb{E}\mathcal{A}_X f_h(Y) - \mathbb{E}\mathcal{A}_Y f_h(Y)\big| \\
&= \frac{1}{N}\bigg|\sum_{s \in [N]} \mathbb{E}\big[(q_X(Y^{(s,1)}|Y) - q_Y(Y^{(s,1)}|Y))\Delta_s f_h(Y)\big]\bigg| \\
&\leq \frac{1}{N} \sum_{s \in [N]} \mathbb{E}\big[\big|q_X(Y^{(s,1)}|Y) - q_Y(Y^{(s,1)}|Y)\big|\big|\Delta_s f_h(Y)\big|\big]. \quad \square
\end{aligned}$$

The next lemma bounds $|\Delta_s f_h(Y)|$.

LEMMA 2.5. *Let* $X, Y \in \{0, 1\}^N$ *be random vectors,* $h : \{0, 1\}^N \to \mathbb{R}$, f_h *be as in Lemma* 2.3. *For each* $s \in [N]$, $x \in \{0, 1\}^N$, *and* $m \geq 0$, *let* $(U^{[x,s]}(m), V^{[x,s]}(m))$ *be any coupling such that* $\mathcal{L}(U^{[x,s]}(m)) = \mathcal{L}(X(m)|X(0) = x^{(s,1)})$ *and* $\mathcal{L}(V^{[x,s]}(m)) = \mathcal{L}(X(m)|X(0) = x^{(s,0)})$. *Then*

$$\big|\Delta_s f_h(x)\big| \leq \sum_{\substack{r \in [N] \\ m \geq 0}} \|\Delta_r h\| \mathbb{P}(U_r^{[x,s]}(m) \neq V_r^{[x,s]}(m)).$$

PROOF. Let $0 = T_0 < T_1 < T_2 < \cdots$ be the jump times of the continuous time Glauber dynamics Markov chain. Since the jump times are independent of the discrete skeleton, and using Lemma 2.3, we have for $x \in \{0, 1\}^N$,

$$\begin{aligned}
\big|\Delta_s f_h(x)\big| &\leq \int_0^\infty \big|\mathbb{E}[h(X(t))|X(0) = x^{(s,1)}] - \mathbb{E}[h(X(t))|X(0) = x^{(s,0)}]\big| \, dt \\
&\leq \sum_{m=0}^\infty \mathbb{E}\bigg|\int_{T_m}^{T_{m+1}} \big[h(U^{[x,s]}(m)) - h(V^{[x,s]}(m))\big] \, dt\bigg| \\
&= \sum_{m=0}^\infty \mathbb{E}\big|h(U^{[x,s]}(m)) - h(V^{[x,s]}(m))\big|
\end{aligned}$$

$$\leq \sum_{\substack{r \in [N] \\ m \geq 0}} \|\Delta_r h\| \mathbb{E} |U_r^{[x,s]}(m) - V_r^{[x,s]}(m)|$$

$$= \sum_{\substack{r \in [N] \\ m \geq 0}} \|\Delta_r h\| \mathbb{P}(U_r^{[x,s]}(m) \neq V_r^{[x,s]}(m)).$$

$\square$

The approach for the proof of Theorem 2.1 is closely related to that of Chatterjee ((2005), Proof of Theorem 4.3).

PROOF OF THEOREM 2.1. Set

$$B := \left(1 - \frac{1}{N}\right) I + \frac{R}{N},$$

where I denotes the identity matrix. We follow the path coupling paradigm of Bubley and Dyer (1997). A direct consequence of Dyer, Goldberg and Jerrum ((2009), Formula (4))—note that their R is the transpose of ours—implies that there is a coupling of the Glauber dynamics of $\mathscr{L}(X)$ such that

$$\mathbb{P}(U_r^{[x,s]}(m) \neq V_r^{[x,s]}(m)) \leq (B^m)_{rs}.$$

Therefore by Lemma 2.5,

$$|\Delta_s f_h(x)| \leq \sum_{\substack{r \in [N] \\ m \geq 0}} \|\Delta_r h\| (B^m)_{rs}.$$

Applying this in Lemma 2.4, and then using Hölder's inequality and sub-multiplicativity of matrix norms, we find

$$|\mathbb{E}h(X) - \mathbb{E}h(Y)| \leq N^{-1} \sum_{\substack{r,s \in [N] \\ m \geq 0}} \mathbb{E}[c_r (B^m)_{rs} v_s(Y)]$$

$$\leq N^{-1} \sum_{m \geq 0} \|c\|_q \mathbb{E} \|B^m v(Y)\|_p$$

$$\leq N^{-1} \|c\|_q \mathbb{E} \|v(Y)\|_p \sum_{m \geq 0} \|B\|_p^m.$$

The result now follows after noting

$$\|B\|_p = \left\|\left(1 - \frac{1}{N}\right) I + \frac{1}{N} R\right\|_p \leq \frac{N-1}{N} + \frac{1}{N} \|R\|_p \leq 1 - \frac{\varepsilon}{N}.$$

$\square$

Specialising the proof of Theorem 2.1 to the case where $p = 1$, Y is a vector of independent Bernoulli variables, and writing X in Gibbs measure form, yields Theorem 1.1, as we now show.

PROOF OF THEOREM 1.1.   Since

$$\mathbb{E}\, d_{\mathrm{H}}\big(\tilde{U}^{[x,s]}, \tilde{V}^{[x,s]}\big) \leq (1-\rho),$$

path coupling implies that for each $s \in [N]$, $x \in \{0,1\}^N$, and $m \geq 0$, there is a coupling $(U^{[x,s]}(m), V^{[x,s]}(m))$ such that $\mathcal{L}(U^{[x,s]}(m)) = \mathcal{L}(X(m)|X(0) = x^{(s,1)})$, and $\mathcal{L}(V^{[x,s]}(m)) = \mathcal{L}(X(m)|X(0) = x^{(s,0)})$, and

$$\mathbb{E}\, d_{\mathrm{H}}\big(U^{[x,s]}(m), V^{[x,s]}(m)\big) \leq (1-\rho)^m.$$

Thus, from Lemma 2.5,

$$\big|\Delta_s f_h(x)\big| \leq \sum_{\substack{r \in [N] \\ m \geq 0}} \|\Delta_r h\| \mathbb{P}\big(U_r^{[x,s]}(m) \neq V_r^{[x,s]}(m)\big)$$

$$\leq \|\Delta h\| \sum_{m \geq 0} \mathbb{E}\, d_{\mathrm{H}}\big(U^{[x,s]}(m), V^{[x,s]}(m)\big)$$

$$= \|\Delta h\| \rho^{-1}.$$

Now, Lemma 2.4 implies

$$\big|\mathbb{E}h(X) - \mathbb{E}h(Y)\big| \leq \frac{\|\Delta h\|}{N\rho} \mathbb{E}\big\|v(Y)\big\|_1,$$

where we use the notation from Theorem 2.1. To bound $\mathbb{E}\|v(Y)\|_1$, first we write the Glauber dynamics probabilities for $\mathscr{L}(X)$ as follows:

$$q_X\big(x^{(s,1)}|x\big) = \mathbb{P}\big(X_s = 1 | (X_u)_{u \neq s} = (x_u)_{u \neq s}\big)$$

$$= \frac{\mathbb{P}(X_s = 1; (X_u)_{u \neq s} = (x_u)_{u \neq s})}{\mathbb{P}(X_s = 1; (X_u)_{u \neq s} = (x_u)_{u \neq s}) + \mathbb{P}(X_s = 0; (X_u)_{u \neq s} = (x_u)_{u \neq s})}$$

$$= \frac{\exp\{L(x^{(s,1)})\}}{\exp\{L(x^{(s,1)})\} + \exp\{L(x^{(s,0)})\}},$$

which simplifies to

$$q_X\big(x^{(s,1)}|x\big) = \frac{e^{\Delta_s L(x)}}{e^{\Delta_s L(x)} + 1} = \frac{1 + \tanh(\frac{1}{2}\Delta_s L(x))}{2}.$$

Now, it is easy to see that

$$(2.3) \qquad p_s = \frac{1 + \tanh(\frac{1}{2}\log(\frac{p_s}{1-p_s}))}{2},$$

and tanh is 1-Lipschitz, so, noting that $q_Y(x^{(s,1)}|x) = p_s$,

$$\big|q_X\big(Y^{(s,1)}|Y\big) - p_s\big| \leq \frac{1}{4}\bigg|\Delta_s L(Y) - \log\bigg(\frac{p_s}{1-p_s}\bigg)\bigg|.$$

A simple rearrangement using (2.3) shows that any solution $p \in [0, 1]^N$ to the system (1.2) also satisfies

$$\log\left(\frac{p_s}{1 - p_s}\right) = \mathbb{E}\Delta_s L(Y),$$

and the result now easily follows. $\square$

3. Proof of ERGM results.   To prove Theorem 1.6, we use Theorem 2.1. In the notation there, we want to bound the influence matrix. The ideas we use essentially come from Bhamidi, Bresler and Sly (2011), where mixing times for the Glauber dynamics of ERGMs are derived. Their results are not strong enough to apply directly, but we are able to derive sharper, more quantitative versions.

For a graph H with v_H vertices and e_H edges, define

$$r_H(x; ij) = \left(\frac{\Delta_{ij}t(H, x)}{2e_H n(n - 1) \cdots (n - v_H + 3)}\right)^{1/(e_H - 1)},$$

and $r_\ell(x; ij) := r_{H_\ell}(x; ij)$. If Z is an Erdős–Rényi graph with parameter a^*, then for any small subgraph H with at least two edges, with high probability, $r_H(Z; ij) = a^* + o(1)$. The basic idea of the proof is that if $r_H(x; ij) \approx a^*$, then the ERGM(β) Glauber dynamics starting from x are similar to the Erdős–Rényi dynamics.

The next lemma makes explicit some statements of Bhamidi, Bresler and Sly ((2011), Proof of Lemma 18). For a graph H and $e \in E(H)$, define the graph $H \setminus e$ to be H with the edge e removed, but with all vertices retained, even if isolated. For a polynomial $f(x) = \sum_{i=1}^{k} a_i x^i$, we use the notation $|f|(x) = \sum_{i=1}^{k} |a_i| x^i$.

LEMMA 3.1.   *For all* $x \in \{0, 1\}^{\binom{n}{2}}$,

$$(3.1) \qquad \sum_{st \neq ij} \left|q\left(x^{(st,1)}|x^{(ij,1)}\right) - q\left(x^{(st,1)}|x^{(ij,0)}\right)\right| \leq \frac{1}{2}|\Phi|'(1).$$

Assuming now $\beta_\ell > 0$ *for* $\ell = 2, \ldots, k$, *if* $a^* \in [0, 1]$ *satisfies* $a^* = \varphi(a^*)$ *and* $x \in \{0, 1\}^{\binom{n}{2}}$ *is such that for all edges* $st \neq ij$ *and all graphs* $H = H_\ell \setminus e$ *for some* $\ell = 2, \ldots, k$ *and* $e \in E(H_\ell)$,

$$(3.2) \qquad \max_{st \neq ij}\left\{\left|r_H\left(x^{(ij,1)}; st\right) - a^*\right| \vee \left|r_H\left(x^{(ij,0)}; st\right) - a^*\right|\right\} \leq \varepsilon,$$

then

$$
\begin{aligned}
&\sum_{st \neq ij} \left|q\left(x^{(st,1)}|x^{(ij,1)}\right) - q\left(x^{(st,1)}|x^{(ij,0)}\right)\right| \\
&\qquad \leq \frac{1}{2}\left[\Phi'(a^*) + \varepsilon\Phi''(1)\right]\left(1 \wedge \left[\operatorname{sech}^2(\Phi(a^*)) + C_2(\varepsilon + n^{-1})\Phi'(1)\right]\right).
\end{aligned}
$$
(3.3)

PROOF.   For $st \neq ij$, we bound

$$|q(x^{(st,1)}|x^{(ij,1)}) - q(x^{(st,1)}|x^{(ij,0)})|$$

$$= \frac{1}{2}\left|\tanh\left(\frac{1}{2}\sum_{\ell=1}^{k}\beta_{\ell}(\Delta_{st}t_{\ell}(x^{(ij,1)}))\right) - \tanh\left(\frac{1}{2}\sum_{\ell=1}^{k}\beta_{\ell}(\Delta_{st}t_{\ell}(x^{(ij,0)}))\right)\right|$$

$$\leq \frac{1}{4}\left(\sum_{\ell=1}^{k}|\beta_{\ell}|(\Delta_{st}\Delta_{ij}t_{\ell}(x))\right)$$

$$\times \left(1 \wedge \left[\text{sech}^2\left(\frac{1}{2}\sum_{\ell=1}^{k}\beta_{\ell}(\Delta_{st}t_{\ell}(x^{(ij,0)}))\right) + \frac{C_2}{4}\left(\sum_{\ell=1}^{k}|\beta_{\ell}|\Delta_{st}\Delta_{ij}t_{\ell}(x)\right)\right]\right),$$

where the inequality follows by using first and second-order Taylor's theorem for tanh. For the first assertion, first we use the 1 in the minimum and sum over edges $st \neq ij$. Next, recall that $\Delta_{st}\Delta_{ij}t(H_{\ell}, x)$ is the number of $(x^{(ij,1)})^{(st,1)}$ edge-preserving injections of $V(H_{\ell})$ into $V(x)$, which use *both* edges ij and st. Writing $t(H_{\ell}, x; e \mapsto st)$ for the number of these injections that map e to st in x, we have

$$\Delta_{st}\Delta_{ij}t(H_{\ell}, x) = \sum_{e\in H_{\ell}}\frac{\Delta_{ij}t(H_{\ell}, x; e \mapsto st)}{n(n-1)\cdots(n-v_{\ell}+3)}$$

and then

(3.4)
$$\sum_{st\neq ij}(\Delta_{st}\Delta_{ij}t_{\ell}(x)) = \sum_{st\neq ij}\sum_{e\in E(H_{\ell})}\frac{\Delta_{ij}t(H_{\ell}, x; e \mapsto st)}{n(n-1)\cdots(n-v_{\ell}+3)}$$

$$= \sum_{e\in H_{\ell}}\frac{\Delta_{ij}t(H_{\ell}\setminus e, x)}{n(n-1)\cdots(n-v_{\ell}+3)},$$

where the first equality is by swapping summations (cf. Bhamidi, Bresler and Sly (2011), Lemma 10). Now, the first assertion follows since

$$\frac{\Delta_{ij}t(H_{\ell}\setminus e, x)}{n(n-1)\cdots(n-v_{\ell}+3)} \leq \frac{\Delta_{ij}t(H_{\ell}\setminus e, \mathbf{1})}{n(n-1)\cdots(n-v_{\ell}+3)} \leq 2(e_{\ell}-1),$$

where $\mathbf{1}$ denotes the complete graph. For the last inequality, note that $\Delta_{ij}t(H_{\ell}\setminus e, \mathbf{1})$ is the number of injections from $V(H_{\ell}\setminus e)$ into $V(\mathbf{1})$ using edge ij. There are at most $2(e_{\ell}-1)$ of assigning vertices of $H_{\ell}\setminus e$ to ij, and there are at most $n(n-1)\cdots(n-v_{\ell}+3)$ ways to assign the remaining vertices. Hence the inequality follows.

For the second assertion, assume that $\beta_{\ell} > 0$ for $\ell = 2, \ldots, k$. For an alternative bound on $\sum_{\ell=1}^{k}\beta_{\ell}(\Delta_{st}\Delta_{ij}t_{\ell}(x))$, note that

$$\Delta_{st}t_{\ell}(x^{(ij,0)}) = 2e_{\ell}r_{\ell}(x^{(ij,0)}; st)^{e_{\ell}-1}.$$

Hence by (3.2) and the fact that; using the positivity of $\beta_2, \ldots, \beta_k$; $0 \le \Phi'(a) \le \Phi'(1) = \sum_{\ell=2}^{k} \beta_\ell e_\ell (e_\ell - 1)$ for $0 < a < 1$, we have

(3.5)
$$
\left| \frac{1}{2} \sum_{\ell=1}^{k} \beta_\ell \left(\Delta_{st} t_\ell \left(x^{(ij,0)} \right) \right) - \Phi(a^*) \right|
$$
$$
= \left| \sum_{\ell=1}^{k} \beta_\ell e_\ell \left(r_\ell \left(x^{(ij,0)}; st \right)^{e_\ell - 1} - (a^*)^{e_\ell - 1} \right) \right|
$$
$$
\le \varepsilon \sum_{\ell=2}^{k} \beta_\ell e_\ell (e_\ell - 1);
$$

we shall derive (3.5), and use it a few more times, below. Then we can write

$$
\operatorname{sech}^2 \left(\frac{1}{2} \sum_{\ell=1}^{k} \beta_\ell \left(\Delta_{st} t_\ell \left(x^{(ij,0)} \right) \right) \right) \le \operatorname{sech}^2 (\Phi(a^*)) + \varepsilon C_2 \sum_{\ell=2}^{k} \beta_\ell e_\ell (e_\ell - 1).
$$

We also bound

(3.6)
$$
\Delta_{st} \Delta_{ij} t_\ell (x) \le \Delta_{st} \Delta_{ij} t_\ell (\mathbf{1}) \le \frac{4 e_\ell (e_\ell - 1)}{n}.
$$

To see (3.5) and (3.6), first note that for $\ell = 2, \ldots, k$, following the same argument as for Remark 1.9, $\Delta_{st} \Delta_{ij} t(H_\ell, x)$ is the number of $(x^{(ij,1)})^{(st,1)}$ edge-preserving injections of $V(H_\ell)$ into $V(x)$, which use *both* edges ij and st. So inequality (3.5) follows from the fact that such subgraph injections increase with the number of edges, maximized for the complete graph $\mathbf{1}$. For inequality (3.6), $\Delta_{st} \Delta_{ij} t(H_\ell, \mathbf{1})$ is simply the number of injections from $V(H_\ell)$ into $V(\mathbf{1})$ using both edges st and ij. To count the number of injections, there are at most $4 e_\ell (e_\ell - 1)$ ways to assign vertices of H_ℓ to st and ij (and maybe fewer depending on the topology of H_ℓ and whether st and ij share a vertex), and then there are at most $(n-3) \cdots (n - v_\ell + 1)$ ways to assign the remaining vertices. The remaining vertices can be counted: $v_\ell - 3$ vertices of H_ℓ to $(n-3)$ remaining vertices of the big graph, when st and ij share a vertex, or $v_\ell - 4$ vertices of H_ℓ to $(n-4)$ remaining vertices otherwise. Thus we find

$$
\Delta_{st} \Delta_{ij} t_\ell (\mathbf{1}) \le 4 e_\ell (e_\ell - 1) \frac{(n-3) \cdots (n - v_\ell + 1)}{n \cdots (n - v_\ell + 3)} \le \frac{4 e_\ell (e_\ell - 1)}{n},
$$

as desired.

Combining the bounds of the previous displays, we have

$$
\begin{aligned}
&\big|q\big(x^{(st,1)}|x^{(ij,1)}\big) - q\big(x^{(st,1)}|x^{(ij,0)}\big)\big| \\
&\quad \leq \left(\sum_{\ell=1}^{k} \beta_\ell\big(\Delta_{st}\Delta_{ij}t_\ell(x)\big)\right) \\
&\qquad \times \left(1 \wedge \left[\frac{\operatorname{sech}^2(\Phi(a^*))}{4} + \left(\varepsilon + \frac{1}{n}\right)\frac{C_2}{4}\sum_{\ell=2}^{k}\beta_\ell e_\ell(e_\ell - 1)\right]\right).
\end{aligned}
$$

Now summing over edges $st \neq ij$, using (3.4), (3.2) and the mean value theorem, we find

$$
\begin{aligned}
\sum_{st \neq ij}\big(\Delta_{st}\Delta_{ij}t_\ell(x)\big) &\leq \sum_{e \in H_\ell}\frac{\Delta_{ij}t(H_\ell \setminus e, x)}{n(n-1)\cdots(n-v_\ell+3)} \\
&\leq 2e_\ell(e_\ell - 1)\big((a^*)^{e_\ell-2} + \varepsilon(e_\ell - 2)\big).
\end{aligned}
$$

We now have

$$
\sum_{\ell=1}^{k}\beta_\ell \sum_{st \neq ij}\big(\Delta_{st}\Delta_{ij}t_\ell(x)\big) \leq 2\Phi'(a^*) + 2\varepsilon\Phi''(1),
$$

and combining the bounds above yields the lemma. □

We prove a quantitative version of Bhamidi, Bresler and Sly ((2011), Lemma 18).

LEMMA 3.2. *If $x \in \{0,1\}^{\binom{n}{2}}$ satisfies the hypotheses of Lemma 3.1, then there is a coupling of two realisations of one step of the embedded discrete time Markov chain for the Glauber dynamics for $\mathrm{ERGM}(\beta)$, denoted $(U^{[x,ij]}(m), V^{[x,ij]}(m))_{m=0,1}$, such that $U^{[x,ij]}(0) = x^{(ij,1)}$, $V^{[x,ij]}(0) = x^{(ij,0)}$ and*

$$
\mathbb{E}\,d_{\mathrm{H}}\big(U^{[x,ij]}(1), V^{[x,ij]}(1)\big) \leq (1 - \rho),
$$

where

$$
\rho = \frac{1 - \frac{1}{2}|\Phi|'(1)}{\binom{n}{2}}.
$$

If $\beta_\ell > 0$ for $\ell = 2, \ldots, k$, then the same inequality holds with

$$
\rho = \frac{1 - \frac{1}{2}[\Phi'(a^*) + \varepsilon\Phi''(1)](1 \wedge [\operatorname{sech}^2(\Phi(a^*)) + C_2(\varepsilon + n^{-1})\Phi'(1)])}{\binom{n}{2}}.
$$

PROOF. At the first step, we choose a common edge at random and update with the optimal coupling of Bernoulli variables, which is unsuccessful for edge st with probability $|q(x^{(st,1)}|x^{(ij,1)}) - q(x^{(st,1)}|x^{(ij,0)})|$. Now note that

$$d_{\mathrm{H}}(U^{[x,ij]}(1), V^{[x,ij]}(1)) - d_{\mathrm{H}}(U^{[x,ij]}(0), V^{[x,ij]}(0))$$

is -1 if edge ij is chosen, is 1 if another edge is chosen and the coupling fails, and is otherwise 0. Thus

$$\mathbb{E} \, d_{\mathrm{H}}(U^{[x,ij]}(1), V^{[x,ij]}(1))$$

$$= 1 - \binom{n}{2}^{-1} + \binom{n}{2}^{-1} \sum_{st \neq ij} |q(x^{(st,1)}|x^{(ij,1)}) - q(x^{(st,1)}|x^{(ij,0)})|,$$

and the result now follows immediately from Lemma 3.1. □

With these lemmas, we can prove Theorem 1.6.

PROOF OF THEOREM 1.6. Note that $0 \leq r_H(y; st) \leq 1$ for any graph H and $y \in \{0, 1\}^{\binom{n}{2}}$, and so setting $\varepsilon = A^*$, we have that (3.2) is automatically satisfied and Lemma 3.2 applies with $\rho = \gamma/\binom{n}{2} > 0$; the positivity by the assumption of the theorem. Applying Theorem 1.1 and using symmetry of the edges proves the result. □

PROOF OF THEOREM 1.7. Lemma 3.2 applies with $\rho = (1 - \frac{1}{2}|\Phi|'(1))/\binom{n}{2} > 0$, so Theorem 1.1 and edge symmetry yields the result. □

PROOF OF PROPOSITION 1.11. Let $k = 2$ and H_2 be a two-star. Applying Theorem 1.7, we compute

$$|\Phi|(a) = |\beta_1| + 2|\beta_2|a, \qquad |\Phi|'(a) = 2|\beta_2|.$$

The condition (1.12) is the same as $|\beta_2| < 1$, and the result now follows after noting $\frac{1}{2}\Delta_{ij}t(H_2, Z) \overset{d}{=} B_1 + B_2$, where $B_1, B_2 \sim \mathrm{Bi}(n-2, a^*)$ and are independent, and so $\mathrm{Var}(\Delta_{ij}t(H_2, Z)) = 8(n-2)a^*(1-a^*)$. □

PROOF OF PROPOSITION 1.12. Let $k = 2$ and H_2 be a triangle. Applying Theorem 1.7, we compute

$$|\Phi|(a) = |\beta_1| + 3|\beta_2|a^2, \qquad |\Phi|'(a) = 6|\beta_2|a,$$

and $|\Phi|'(1) = 6|\beta_2|$. Then we can apply Theorem 1.7 assuming (1.12), which is the same as $|\beta_2| < 1/3$. The result now follows after noting that $\frac{1}{6}\Delta_{ij}t(H_2, Z) \sim \mathrm{Bi}(n-2, (a^*)^2)$ and so $\mathrm{Var}(\Delta_{ij}t(H_2, Z)) = 36(n-2)(a^*)^2(1-(a^*)^2)$. □

PROOF OF THEOREM 1.13. As Bhamidi, Bresler and Sly (2011) showed, the ERGM chain may not be contracting over the entire state space in the high temperature regime. Thus we will not use Theorem 2.1, but instead combine and apply Lemmas 2.4 and 2.5 directly.

For each $m \geq 0$, let $(U^{[x,ij]}(m), V^{[x,ij]}(m))$ be a coupling of the Glauber dynamics of the ERGM(β), with initial states $U^{[x,ij]}(0) = x^{(ij,1)}$, $V^{[x,ij]}(0) = x^{(ij,0)}$. Lemma 2.5 implies that

$$\left| \Delta_{ij} f_h(x) \right| \leq \|\Delta h\| \sum_{\substack{st \in [n]_2 \\ m \geq 0}} \mathbb{P}(U_{st}^{[x,ij]}(m) \neq V_{st}^{[x,ij]}(m))$$

$$(3.7) \qquad \leq \|\Delta h\| \left[\sum_{m < t_n} \mathbb{E} \, d_{\mathrm{H}}(U^{[x,ij]}(m), V^{[x,ij]}(m)) \right.$$

$$(3.8) \qquad \left. + \binom{n}{2} \sum_{m \geq t_n}^{\infty} \mathbb{P}(U^{[x,ij]}(m) \neq V^{[x,ij]}(m)) \right],$$

where we set $t_n := cn^2 \log(n)^2$, with c to be chosen large enough (for (3.9) below). The main idea of the proof is to bound (3.8) by using that the chain mixes in $n^2 \log(n)$ steps, and to bound (3.7) separately for the two cases where the x is such that the chain stays in a contracting region for t_n steps with high probability, and where x is not. Then we show the former case occurs with high probability under the Erdős–Rényi distribution.

To bound (3.8), we use Bhamidi, Bresler and Sly ((2011), Theorem 5), which states that the $1/4$-mixing time t_{mix} of these Glauber dynamics is of order $t_{\mathrm{mix}} = \mathrm{O}(n^2 \log(n))$ and thus, using Levin, Peres and Wilmer ((2009), Exercise 4.3 and Formula (4.33)), there are couplings for $m \geq t_n$, such that for some $d > 0$,

$$\mathbb{P}(U^{[x,ij]}(m) \neq V^{[x,ij]}(m)) \leq \mathbb{P}\left(U^{[x,ij]}\left(\left\lfloor \frac{m}{t_{\mathrm{mix}}} \right\rfloor t_{\mathrm{mix}} \right) \neq V^{[x,ij]}\left(\left\lfloor \frac{m}{t_{\mathrm{mix}}} \right\rfloor t_{\mathrm{mix}} \right) \right)$$

$$\leq 2^{-\lfloor \frac{m}{t_{\mathrm{mix}}} \rfloor}$$

$$\leq \left(e^{-\frac{d}{n^2 \log(n)}} \right)^m.$$

We now easily find that for a constant d' not depending on c,

$$(3.9) \qquad \binom{n}{2} \sum_{m \geq t_n}^{\infty} \mathbb{P}(U^{[x,ij]}(m) \neq V^{[x,ij]}(m)) \leq d' n^4 \log(n) n^{-dc}.$$

Now, assume n is large enough so that there is an $\varepsilon > 0$ such that (3.3) is less than one; call the right-hand side of (3.3) $\alpha < 1$. This is always possible since (3.3) decreases as both $\varepsilon \to 0$ and $n \to \infty$, and has limit $\varphi'(a^*)$, which is strictly less than one by assumption. For $\nu > 0$, let $\mathcal{E}_\nu$ be the set of $x \in \{0, 1\}^{\binom{n}{2}}$ such that for all $ij \in [n]_2$ and H as in Lemma 3.1, (3.2) with ε replaced by ν is satisfied. We only consider $\nu \leq \varepsilon$, so that the chain is contracting while in $\mathcal{E}_\nu$.

Let $0 \le m < t_n$, and $(U^{[x,ij]}(m), V^{[x,ij]}(m))$ be the coupling from Bhamidi, Bresler and Sly (2011). Denote the event that the coupling (and hence the path coupling by monotonicity) stays in the region $\mathcal{E}_\nu$ up to time t_n by

$$B_\nu := \bigcap_{m < t_n} \{U^{[x,ij]}(m) \in \mathcal{E}_\nu, V^{[x,ij]}(m) \in \mathcal{E}_\nu\}.$$

Combining (18) of Bhamidi, Bresler and Sly (2011) and the comment before Claim 15 of Bhamidi, Bresler and Sly (2011), as well as the comment before Lemma 16 of Bhamidi, Bresler and Sly (2011), we find that for any $\mu > 0$, $\delta > 0$ small enough, if $x \in \mathcal{E}_{2\mu} \setminus \mathcal{E}_\mu$, then there is a $d'' > 0$ such that

(3.10) $$\mathbb{P}(B^c_{2\mu+\delta}) \le t_n^4 e^{-d''n}.$$

If $x \in \mathcal{E}_\mu$, then in order for either chain to leave $\mathcal{E}_{2\mu+\delta}$ it must first enter $\mathcal{E}_{2\mu} \setminus \mathcal{E}_\mu$, and so the Markov property and the argument above implies that (3.10) holds also for $x \in \mathcal{E}_\mu$. Now, choosing μ, δ so that $2\mu + \delta < \varepsilon$ and, in particular, $B^c_\varepsilon \subseteq B^c_{2\mu+\delta}$, with Lemma 3.1 we find that for $m < t_n$,

$$\mathbb{E}\, d_{\mathrm{H}}\big(U^{[x,ij]}(m), V^{[x,ij]}(m)\big) \le \mathbb{I}[x \in \mathcal{E}_\mu]\mathbb{P}[B_\varepsilon]\left(1 - (1-\alpha)\binom{n}{2}^{-1}\right)^m$$

$$+ \big(\mathbb{I}[x \notin \mathcal{E}_\mu] + \mathbb{P}[B^c_\varepsilon]\big)\binom{n}{2}.$$

Combining these last two displays with (3.10), (3.9), (3.8), (3.7), noting the chain is contracting under B_ε, and choosing c large enough yields

$$|\Delta_{ij} f_h(x)| \le \|\Delta h\|\binom{n}{2}\big(\{(1-\alpha)^{-1} + o(1)\}\mathbb{I}[x \in \mathcal{E}_\mu] + t_n\mathbb{I}[x \notin \mathcal{E}_\mu]\big).$$

Now applying Lemma 2.4 and noting that $|q_X(\cdot|y) - q_Y(\cdot|y)| \le 1$, we have

$$\big|\mathbb{E}h(X) - \mathbb{E}h(Z)\big|$$

$$\le \|\Delta h\|\bigg\{((1-\alpha)^{-1} + o(1))$$

$$\times \sum_{ij \in [n]_2} \mathbb{E}\big[\mathbb{I}[Z \in \mathcal{E}_\mu]\big|q_X\big(Z^{(ij,1)}|Z\big) - q_Z\big(Z^{(ij,1)}|Z\big)\big|\big]$$

$$+ t_n \sum_{ij \in [n]_2} \mathbb{P}(Z \notin \mathcal{E}_\mu)\bigg\}.$$

The first term (after dropping the indicator) is bounded the same as in the proof of Theorem 1.6 and Corollary 1.10. For the second term, we show that for all H under consideration and $st \ne ij$, both $r_H(Z^{(ij,1)}, st)$ and $r_H(Z^{(ij,0)}, st)$ are highly concentrated, and then the result follows from a union bound.

From the definition of r_H and elementary considerations of the function of the function $[0, 1] \ni w \mapsto w^{1/(e_H - 1)}$, we have

$$\mathbb{P}\big(\big|r_H\big(Z^{(ij,\cdot)}, st\big) - a^*\big| > \mu\big)$$

$$\leq \mathbb{P}\bigg(\bigg|\frac{\Delta_{st}t(H, Z^{(ij,\cdot)})}{2e_H n(n-1)\cdots(n-v_H+3)} - (a^*)^{e_H-1}\bigg| > \mu^{e_H-1}\bigg).$$

Since

$$\frac{\Delta_{st}t(H, Z^{(ij,\cdot)})}{2e_H n(n-1)\cdots(n-v_H+3)} = \frac{\Delta_{st}t(H, Z)}{2e_H n(n-1)\cdots(n-v_H+3)} + \mathrm{O}(n^{-1})$$

$$= \frac{\Delta_{st}t(H, Z)}{2e_H (n-2)\cdots(n-v_H+1)} + \mathrm{O}(n^{-1}),$$

we instead bound

$$(3.11) \qquad \mathbb{P}\bigg(\bigg|\frac{\Delta_{st}t(H, Z)}{2e_H (n-2)\cdots(n-v_H+1)} - (a^*)^{e_H-1}\bigg| > \mu'\bigg),$$

for sufficiently large n and small $\mu' > 0$. Using the argument of Ruciński ((1988), equation (5), page 6); see also Remark 1.9; we have that for fixed a^*, and fixed $\ell \geq 1$,

$$\mathbb{E}\big[\big(\Delta_{st}t(H, Z) - \mathbb{E}\Delta_{st}t(H, Z)\big)^{2\ell}\big] = \mathrm{O}\big(n^{\ell(2|v(H)|-5)}\big).$$

Now noting that

$$\mathbb{E}\Delta_{st}t(H, Z) = 2e_H (n-2)\cdots(n-v_H+1)(a^*)^{e_H-1},$$

we use Markov's inequality in (3.11) to find that for fixed μ',

$$\mathbb{P}\bigg(\bigg|\frac{\Delta_{st}t(H, Z)}{2e_H (n-2)\cdots(n-v_H+1)} - (a^*)^{e_H-1}\bigg| > \mu'\bigg) = \mathrm{O}(n^{-\ell}).$$

By choosing ℓ large enough and using a union bound, we have that $\mathbb{P}(Z \notin \mathcal{E}_\mu) = \mathrm{O}(n^{-1/2})$, as desired. $\square$

REFERENCES

ALI, W., RITO, T., REINERT, G., SUN, F. and DEANE, C. M. (2014). Alignment-free protein interaction network comparison. *Bioinformatics* **30** i430–i437.

BARBOUR, A. D. (1988). Stein's method and Poisson process convergence. A celebration of applied probability. *J. Appl. Probab.* **25** 175–184. MR0974580

BARBOUR, A. D. (1990). Stein's method for diffusion approximations. *Probab. Theory Related Fields* **84** 297–322. MR1035659

BARBOUR, A. D. and CHEN, L. H. Y., eds. (2005). *An Introduction to Stein's Method. Lecture Notes Series. Institute for Mathematical Sciences. National University of Singapore* **4**. World Scientific, Hackensack, NJ. MR2235447

BARBOUR, A. D., RÖLLIN, A. and ROSS, N. (2019). Error bounds in local limit theorems using Stein's method. *Bernoulli* **25** 1076–1104. MR3920366

BHAMIDI, S., BRESLER, G. and SLY, A. (2011). Mixing time of exponential random graphs. *Ann. Appl. Probab.* **21** 2146–2170. MR2895412

BORGS, C., CHAYES, J. T., LOVÁSZ, L., SÓS, V. T. and VESZTERGOMBI, K. (2008). Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.* **219** 1801–1851. MR2455626

BORGS, C., CHAYES, J. T., LOVÁSZ, L., SÓS, V. T. and VESZTERGOMBI, K. (2012). Convergent sequences of dense graphs II. Multiway cuts and statistical physics. *Ann. of Math.* (2) **176** 151–219. MR2925382

BRESLER, G. and NAGARAJ, D. (2017). Stein's method for stationary distributions of Markov chains and application to Ising models. Preprint. Available at arXiv:1712.05743.

BUBLEY, R. and DYER, M. (1997). Path coupling: A technique for proving rapid mixing in Markov chains. In *Proceedings 38th Annual Symposium on Foundations of Computer Science* 223–231.

CHATTERJEE, S. (2005). *Concentration Inequalities with Exchangeable Pairs*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)—Stanford University. Available at arXiv:math/0507526. MR2707160

CHATTERJEE, S. (2007). Stein's method for concentration inequalities. *Probab. Theory Related Fields* **138** 305–321. MR2288072

CHATTERJEE, S. (2016). An introduction to large deviations for random graphs. *Bull. Amer. Math. Soc.* (*N.S.*) **53** 617–642. MR3544262

CHATTERJEE, S. and DEMBO, A. (2016). Nonlinear large deviations. *Adv. Math.* **299** 396–450. MR3519474

CHATTERJEE, S. and DIACONIS, P. (2013). Estimating and understanding exponential random graph models. *Ann. Statist.* **41** 2428–2461. MR3127871

CHATTERJEE, S. and SHAO, Q.-M. (2011). Nonnormal approximation by Stein's method of exchangeable pairs with application to the Curie–Weiss model. *Ann. Appl. Probab.* **21** 464–483. MR2807964

CHEN, L. H. Y., GOLDSTEIN, L. and SHAO, Q.-M. (2011). *Normal Approximation by Stein's Method. Probability and Its Applications* (*New York*). Springer, Heidelberg. MR2732624

DOBRUSHIN, R. L. and SHLOSMAN, S. B. (1985). Constructive criterion for the uniqueness of Gibbs field. In *Statistical Physics and Dynamical Systems* (*Köszeg*, 1984). *Progress in Probability* **10** 347–370. Birkhäuser, Boston, MA. MR0821306

DOBRUSHIN, R. L. (1970). Definition of a system of random variables by means of conditional distributions. *Theory Probab. Appl.* **15** 458–486.

DYER, M., GOLDBERG, L. A. and JERRUM, M. (2009). Matrix norms and rapid mixing for spin systems. *Ann. Appl. Probab.* **19** 71–107. MR2498672

EICHELSBACHER, P. and LÖWE, M. (2010). Stein's method for dependent random variables occurring in statistical mechanics. *Electron. J. Probab.* **15** 962–988. MR2659754

EICHELSBACHER, P. and REINERT, G. (2008). Stein's method for discrete Gibbs measures. *Ann. Appl. Probab.* **18** 1588–1618. MR2434182

ELDAN, R. (2018). Gaussian-width gradient complexity, reverse log-Sobolev inequalities and nonlinear large deviations. *Geom. Funct. Anal.* **28** 1548–1596. MR3881829

ELDAN, R. and GROSS, R. (2018a). Decomposition of mean-field Gibbs distributions into product measures. *Electron. J. Probab.* **23** Paper No. 35, 24. MR3798245

ELDAN, R. and GROSS, R. (2018b). Exponential random graphs behave like mixtures of stochastic block models. *Ann. Appl. Probab.* **28** 3698–3735. MR3861824

ELLIS, R. S. (2006). *Entropy, Large Deviations, and Statistical Mechanics. Classics in Mathematics.* Springer, Berlin. MR2189669

ELLIS, R. S. and NEWMAN, C. M. (1978). Limit theorems for sums of dependent random variables occurring in statistical mechanics. *Z. Wahrsch. Verw. Gebiete* **44** 117–139. MR0503333

ELLIS, R. S., NEWMAN, C. M. and ROSEN, J. S. (1980). Limit theorems for sums of dependent random variables occurring in statistical mechanics. II. Conditioning, multiple phases, and metastability. *Z. Wahrsch. Verw. Gebiete* **51** 153–169. MR0566313

FRANK, O. and STRAUSS, D. (1986). Markov graphs. *J. Amer. Statist. Assoc.* **81** 832–842. MR0860518

GHEISSARI, R., LUBETZKY, E. and PERES, Y. (2018). Concentration inequalities for polynomials of contracting Ising models. *Electron. Commun. Probab.* **23** Paper No. 76, 12. MR3873783

HAYES, T. P. (2006). A simple condition implying rapid mixing of single-site dynamics on spin systems. In *Proc. 47th Annual IEEE Symposium on Foundations of Computer Science FOCS '06* 39–46. IEEE Computer Society, Washington, DC.

HOLLAND, P. W. and LEINHARDT, S. (1981). An exponential family of probability distributions for directed graphs. *J. Amer. Statist. Assoc.* **76** 33–65. MR0608176

LEVIN, D. A., LUCZAK, M. J. and PERES, Y. (2010). Glauber dynamics for the mean-field Ising model: Cut-off, critical power law, and metastability. *Probab. Theory Related Fields* **146** 223–265. MR2550363

LEVIN, D. A., PERES, Y. and WILMER, E. L. (2009). *Markov Chains and Mixing Times.* Amer. Math. Soc., Providence, RI. MR2466937

LEY, C., REINERT, G. and SWAN, Y. (2017). Stein's method for comparison of univariate distributions. *Probab. Surv.* **14** 1–52. MR3595350

LUCZAK, M. J. (2008). Concentration of measure and mixing for Markov chains. In *Fifth Colloquium on Mathematics and Computer Science. Discrete Math. Theor. Comput. Sci. Proc.*, AI 95–120. Assoc. Discrete Math. Theor. Comput. Sci., Nancy. MR2508781

MILO, R., SHEN-ORR, S., ITZKOVITZ, S., KASHTAN, N., CHKLOVSKII, D. and ALON, U. (2002). Network motifs: Simple building blocks of complex networks. *Science* **298** 824–827.

MOSSEL, E. and SLY, A. (2013). Exact thresholds for Ising–Gibbs samplers on general graphs. *Ann. Probab.* **41** 294–328. MR3059200

PADGETT, J. F. and ANSELL, C. K. (1993). Robust action and the rise of the Medici, 1400–1434. *Amer. J. Sociol.* **98** 1259–1319.

PICARD, F., DAUDIN, J.-J., KOSKAS, M., SCHBATH, S. and ROBIN, S. (2008). Assessing the exceptionality of network motifs. *J. Comput. Biol.* **15** 1–20. MR2383618

PRŽULJ, N. (2007). Biological network comparison using graphlet degree distribution. *Bioinformatics* **23** e177–e183.

RADIN, C. and YIN, M. (2013). Phase transitions in exponential random graphs. *Ann. Appl. Probab.* **23** 2458–2471. MR3127941

REINERT, G. (2005). Three general approaches to Stein's method. In *An Introduction to Stein's Method. Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.* **4** 183–221. Singapore Univ. Press, Singapore. MR2235451

RINOTT, Y. and ROTAR, V. (1997). On coupling constructions and rates in the CLT for dependent summands with applications to the antivoter model and weighted *U*-statistics. *Ann. Appl. Probab.* **7** 1080–1105. MR1484798

RÖLLIN, A. and ROSS, N. (2015). Local limit theorems via Landau–Kolmogorov inequalities. *Bernoulli* **21** 851–880. MR3338649

ROSS, N. (2011). Fundamentals of Stein's method. *Probab. Surv.* **8** 210–293. MR2861132

RUCIŃSKI, A. (1988). When are small subgraphs of a random graph normally distributed? *Probab. Theory Related Fields* **78** 1–10. MR0940863

SINULIS, A. (2018). Mixing times of Glauber dynamics via entropy methods. Preprint. Available at arXiv:1804.04424.

STEIN, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* (*Univ. California*, *Berkeley*, *Calif.*, 1970/1971), *Vol. II*: *Probability Theory* 583–602. Univ. California Press, Berkeley, CA. MR0402873

STEIN, C. (1986). *Approximate Computation of Expectations*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **7**. IMS, Hayward, CA. MR0882007

WASSERMAN, S. and FAUST, K. (1994). *Social Network Analysis*: *Methods and Applications*. Cambridge Univ. Press, Cambridge.

YIN, M. (2013). Critical phenomena in exponential random graphs. *J. Stat. Phys.* **153** 1008–1021. MR3131507

DEPARTMENT OF STATISTICS
UNIVERSITY OF OXFORD
24-29 ST GILES'
OXFORD, OX1 3LB
UNITED KINGDOM
E-MAIL: reinert@stats.ox.ac.uk

SCHOOL OF MATHEMATICS AND STATISTICS
UNIVERSITY OF MELBOURNE
PETER HALL BUILDING
MELBOURNE, VICTORIA, 3010
AUSTRALIA
E-MAIL: nathan.ross@unimelb.edu.au