

MIXING TIMES OF GLAUBER DYNAMICS VIA ENTROPY METHODS

A. SINULIS

ABSTRACT. In this work we prove sufficient conditions for the Glauber dynamics corresponding to a sequence of (non-product) measures on finite product spaces to be rapidly mixing, i.e. that the mixing time with respect to the total variation distance satisfies $t_{mix} = O(N \log N)$, where N is the system size. The proofs do not rely on coupling arguments, but instead use functional inequalities. As a byproduct, we obtain exponential decay of the relative entropy along the Glauber semigroup.

These conditions can be checked in various examples, which include the exponential random graph models with sufficiently small parameters (which does not require any monotonicity in the system and thus also applies to negative parameters, as long the associated monotone system is in the high temperature phase), the vertex-weighted exponential random graph models, as well as models with hard constraints such as the random coloring and the hard-core model.

1. INTRODUCTION

Spin systems are ubiquitous in the modeling of various phenomenons, ranging from toy models to explain ferromagnetism (the Ising and the Potts model, or more generally the random cluster model), to voter models, various network models (such as the Erdős-Renyi model or the exponential random graph models) and models with *hard constraints* such as the random proper coloring model or the hard-core model.

Spin systems can be described as probability measures on finite product spaces, and hard constraints translate into conditions on the support of the probability measure. A popular approach to define a spin system is by specifying a Hamiltonian function H defined on the space of *configurations* and set $\mu(x) = Z^{-1} \exp(H(x))$, where x is a configuration. Informally, hard constraints can be incorporated by setting $H(y) = -\infty$ for a non-admissible configuration. More formally, we will consider a finite set \mathcal{X} (the spins), a finite set \mathcal{I} (the sites) and the spin system is a measure μ on $\mathcal{Y} := \mathcal{X}^{\mathcal{I}}$, and we are interested in the mixing time asymptotic of the Glauber dynamics on a sequence of spin systems.

1.1. Mixing times and the Glauber dynamics. It is often important to *sample* from the spin system under consideration. In most cases, however, the normalization constant $Z = \sum_{\sigma \in \mathcal{Y}} \mu(\sigma)$ cannot be computed efficiently, as the number of sites increases. It is necessary to bypass this problem; one way is to construct a Markov chain converging to the spin system, and evaluating the time to stationary becomes crucial as the size of the system grows.

Date: July 21, 2018.

1991 Mathematics Subject Classification. Primary 60J, Secondary 82B.

Key words and phrases. Glauber dynamics, logarithmic Sobolev inequalities, mixing times, spin systems.

This research was supported by CRC 1283.

One choice is to use the associated Glauber dynamics of the spin system, which is a \mathcal{Y} -valued ergodic Markov chain $(Y_t)_{t \in \mathbb{N}_0}$ with reversible (and thus stationary) distribution μ . It is known that under mild assumptions the distribution of $(Y_t)_t$ will converge to the stationary distribution. At each step, the Glauber dynamics selects a site $i \in \mathcal{I}$ uniformly at random and updates it with the conditional probability given \bar{x}_i , i.e. its transition probability is given by

$$P(x, y) = |\mathcal{I}|^{-1} \mu(y_i \mid \bar{x}_i) \mathbb{1}_{x_j=y_j \ \forall j \neq i}.$$

Here $\mathbb{1}_A$ is the indicator function of the event A .

Now if $(Y_t)_{t \in \mathbb{N}_0}$ is a Markov chain on any finite space \mathcal{Y} with a reversible measure ν , this convergence can be quantified by using various metrics between probability measures. One canonical way is to choose the total variation distance

$$(1.1) \quad d_{TV}(\mu_1, \mu_2) := \sup_{A \subset \mathcal{Y}} |\mu_1(A) - \mu_2(A)| = \frac{1}{2} \sum_{x \in \mathcal{Y}} |\mu_1(x) - \mu_2(x)|$$

to define the mixing time

$$(1.2) \quad t_{mix} := \inf \{ t \in \mathbb{N}_0 : \sup_{y \in \mathcal{Y}} d_{TV}(\delta_y * P^t, \nu) \leq e^{-1} \},$$

or for any \mathcal{Y} -valued Markov process $(Y_t)_{t \in \mathbb{R}_+}$

$$(1.3) \quad t_{mix} := \inf \{ t \in \mathbb{R}_+ : \sup_{y \in \mathcal{Y}} d_{TV}(\delta_y * P^t, \nu) \leq e^{-1} \}.$$

Here, we denote by $\delta_y * P^t$ the distribution of Y_t given that the Markov chain starts at y . We shall mainly work with the continuous-time version of the Glauber dynamics, and thus use (1.3). Another, maybe less canonical, way to quantify the speed of convergence is to use the relative entropy between two measures μ, ν on any measurable space defined as

$$H(\mu \parallel \nu) = \begin{cases} \int \frac{d\mu}{d\nu} \log \left(\frac{d\mu}{d\nu} \right) d\nu & \mu \ll \nu \\ 0 & \text{otherwise,} \end{cases}$$

and we can define the mixing time $t_{mix,ent}$ as above, replacing $d_{TV}(\delta_y * P^t, \nu)$ by $H(\delta_y * P^t \parallel \nu)$.

1.2. Functional inequalities and tensorization of entropy. In the context of concentration of measure, functional inequalities have become prominent and important in the nineties, since these yielded easier proofs of known (and previously unknown) concentration results. For an introduction to the concentration of measure phenomenon and functional inequalities we refer to [Led01] or more recently [BLM13]. P. Diaconis and L. Saloff-Coste used functional inequalities, especially logarithmic Sobolev inequalities, to obtain mixing times of various Markov chains in [DS96]. Moreover, by works of M. Ledoux and S. G. Bobkov different notions of so-called modified logarithmic Sobolev inequalities have been paid attention to, see [BL98; GQ03] and the work by S. G. Bobkov and P. Tetali [BT06].

Let us give a slight exposition into functional inequalities in the framework of Markov chains. Let \mathcal{Y} be a finite set, P be the transition matrix of a Markov chain on \mathcal{Y} and $-L = I - P$ be its generator. If P is reversible with respect to a measure μ , we can define the entropy functional

$$(1.4) \quad \text{Ent}_\mu(f) := \mathbb{E}_\mu f \log f - \mathbb{E}_\mu f \log(\mathbb{E}_\mu f) \quad \text{for } f \geq 0$$

and the Dirichlet form

$$(1.5) \quad \mathcal{E}(f, g) := -\mathbb{E}_\mu(fLg).$$

We say that the triple (\mathcal{Y}, P, μ) (or in short P , if the space and the measure are clear from the context) satisfies a *logarithmic Sobolev inequality* with constant ρ if for all $f : \mathcal{Y} \rightarrow \mathbb{R}$

$$(1.6) \quad \text{Ent}_\mu(f^2) \leq 2\rho\mathcal{E}(f, f),$$

and that it satisfies a *modified logarithmic Sobolev inequality* with constant ρ_0 , if for all $f : \mathcal{Y} \rightarrow \mathbb{R}_+$ we have

$$(1.7) \quad \text{Ent}_\mu(f) \leq \frac{\rho_0}{2}\mathcal{E}(f, \log f).$$

The best constant in (1.6) ((1.7) respectively) is known as (modified) logarithmic Sobolev constant, cf. [BT06, equations (1.5) and (1.7)], where our constants ρ, ρ_0 correspond to their constants $1/\rho, 1/\rho_0$. The modified logarithmic Sobolev constant is also called entropy constant, see e.g. the definition of β in [GQ03]. It is known that the modified logarithmic Sobolev constant can be used to bound mixing time for the total variation distance of (the distribution of) a Markov semigroup and its trend to equilibrium, and sometimes gives sharper results than using the logarithmic Sobolev constant (in the sense of Gross, [Gro75]).

To establish the connection between modified logarithmic Sobolev inequalities and the mixing time of the continuous-time Markov process with generator L , let us state a Theorem (and Corollary) by S. G. Bobkov and P. Tetali, see [BT06, Theorem 2.4, Corollary 2.8]. Note that our logarithmic Sobolev constant ρ_0 corresponds to $1/\rho_0$ in [BT06].

Theorem 1.1 (Bobkov-Tetali). *Let μ_0 be any measure on a finite set \mathcal{Y} and denote by μ_t the distribution of the Markov process $(X_t)_t$ with initial distribution μ_0 and generator L and by f_t its density with respect to the reversible measure π . Then for any $t \geq 0$*

$$H(\mu_t \parallel \pi) = \text{Ent}_\pi(f_t) \leq H(\mu_0 \parallel \pi)e^{-\frac{2}{\rho_0}t},$$

and consequently

$$d_{TV}(\mu_t, \pi)^2 \leq 2H(\mu_t \parallel \pi) \leq 2\log\left(\frac{1}{\pi^*}\right)e^{-\frac{2}{\rho_0}t},$$

where $\pi^* := \min_{x \in \mathcal{Y}} \pi(x)$

Moreover, we shall require a powerful tool in the framework of product spaces, namely the tensorization property of the (modified) logarithmic Sobolev inequality. Since we are working with a non-product measure (and thus the individual spins are not independent), we need the concept of weakly dependent random variables. Let μ a spin system on $\mathcal{Y} = \mathcal{X}^{\mathcal{I}}$, and define an *interdependence matrix* $(J_{ij})_{i,j \in \mathcal{I}}$ as any matrix with $J_{ii} = 0$ and such that for any $x, y \in \mathcal{Y}$ differing only in the j -th site we have

$$d_{TV}(\mu(\cdot \mid \bar{x}_i), \mu(\cdot \mid \bar{y}_i)) \leq J_{ij}.$$

By $\mu(\cdot \mid \bar{x}_i)$ we always mean the conditional probability, interpreted as a measure on \mathcal{X} . Note that if μ is a product measure, then $J \equiv 0$ is an interdependence matrix, and thus J (or any norms thereof) measures the strength of interaction between the spins in the spin system μ .

We will need the following approximate tensorization result of the entropy initially proven by K. Marton [Mar15] (see also [GSS18, Theorem 4.1]), on which the proof of Theorem 1.3 is based. For the reader's convenience, we shall formulate it in our setting.

Theorem 1.2 (Marton). *Let μ be a measure on a product space $\mathcal{Y} := \mathcal{X}^{\mathcal{I}}$ for some finite sets \mathcal{X} and \mathcal{I} . If for some $\alpha_1, \alpha_2 > 0$*

$$\tilde{\beta}(\mu) := \inf_{S \subset \mathcal{I}} \inf_{i \notin S} \tilde{\beta}_{i,S}(\mu) \geq \alpha_1 > 0$$

where

$$\tilde{\beta}_{i,S}(\mu) := \inf_{\substack{x^S \in \mathcal{X}^S \\ \mu(x^S) > 0}} \inf_{\substack{y^{S^c} \in \mathcal{X}^{S^c} \\ \mu(y^{S^c}, x^S) > 0}} \mu_{\bar{S}}((y^{S^c})_i | x^S)$$

holds and an interdependence matrix J satisfies $\|J\|_{2 \rightarrow 2} \leq 1 - \alpha_2$, then for any function $f : \mathcal{Y} \rightarrow \mathbb{R}_+$ vanishing outside of $\text{supp } \mu$ we have

$$(1.8) \quad \text{Ent}_{\mu}(f) \leq \frac{1}{\alpha_1 \alpha_2} \sum_{i \in \mathcal{I}} \int \text{Ent}_{\mu(\cdot | \bar{x}_i)}(f(\bar{x}_i, \cdot)) d\mu(x).$$

We will not give a proof here, but only note that the inductive approach given in [Mar15] (or see [GSS18, Theorem 4.1]) also works in the case of μ not having full support (i.e. the spin system having hard constraints) since α_1 is a uniform lower bound for any subset $S \subset \mathcal{I}$, any $x \in \mathcal{X}^S$ with $\mu_S(x) > 0$ and any $i \notin S$. In the first infimum, the choice $S = \emptyset$ is considered as well, which has to be read as $\tilde{\beta}_{i,\emptyset}(\mu) = \inf_{y \in \mathcal{Y} : \mu(y) > 0} \mu(y_i)$. The interpretation of $\tilde{\beta}_{i,S}(\mu)$ is straightforward: For any admissible partial configuration $x_S \in \mathcal{X}^S$ all possible marginals are supported on points with probability at least α_1 .

If there are no hard constraints, i.e. μ has full support, then $\tilde{\beta}(\mu)$ can be simplified to

$$\tilde{\beta}(\mu) = I(\mu) := \min_{i \in \mathcal{I}} \min_{y \in \mathcal{Y}} \mu_n(y_i | \bar{y}_i),$$

which can be shown by conditioning for any $S \subset \mathcal{I}$ and any $x_S \in \mathcal{X}^S$ as follows

$$\mu(y_i | x_S) = \mu(x_S)^{-1} \sum_{z \in \mathcal{X}^{\mathcal{I} \setminus (S \cup i)}} \mu(y_i | x_S, z) \mu(x_S, z) \geq I(\mu),$$

and the reverse inequality follows by taking $S = \mathcal{I} \setminus \{j\}$.

1.3. Main result. We are now ready to state our main result on the mixing time of Glauber dynamics associated to spin systems.

Theorem 1.3. *Let \mathcal{X}, \mathcal{I} be finite sets, $\mathcal{Y} := \mathcal{X}^{\mathcal{I}}$ and μ be a measure on \mathcal{Y} . Assume that for some constants $\alpha_1, \alpha_2 > 0$ we have the lower bound on the conditional probabilities*

$$(1.9) \quad \tilde{\beta}(\mu) \geq \alpha_1$$

and an upper bound on the interdependence matrix J (also known as Dobrushin's uniqueness condition)

$$(1.10) \quad \|J\|_{2 \rightarrow 2} \leq 1 - \alpha_2.$$

The Glauber dynamics associated to μ satisfies a modified logarithmic Sobolev inequality with constant $2|\mathcal{I}|\alpha_1^{-1}\alpha_2^{-2}$. As a consequence, given any initial distribution $\mu_0 = f_0\mu$ on \mathcal{Y} , the distribution μ_t of $(X_t)_t$ satisfies

$$(1.11) \quad H(\mu_t \parallel \mu) \leq H(\mu_0 \parallel \mu) \exp\left(-\frac{\alpha_1\alpha_2^2}{2|\mathcal{I}|}t\right).$$

Furthermore, if $(\mu_n)_n$ is a sequence of spin systems with sites $(\mathcal{I}_n)_n$ satisfying (1.9) and (1.10) uniformly, then the sequence of Glauber dynamics is rapidly mixing, i.e. $t_{mix} = O(|\mathcal{I}_n| \log|\mathcal{I}_n|)$.

In the case of spin systems without hard constraints, we can rephrase the conditions.

Corollary 1.4. *Let $(\mu_n)_n$ be a sequence of Gibbs measures on configuration spaces \mathcal{Y}_n , i.e. for some Hamiltonian $H_n : \mathcal{Y}_n \rightarrow \mathbb{R}$ we have*

$$(1.12) \quad \mu_n(y) = Z_n^{-1} \exp(H_n(y)).$$

If

$$(1.13) \quad I(\mu_n) \geq \alpha_1$$

$$(1.14) \quad \|\mathcal{J}_n\|_{2 \rightarrow 2} \leq 1 - \alpha_2$$

for some $\alpha_1, \alpha_2, C > 0$, then the (sequence of) Glauber dynamics associated to μ_n is rapidly mixing.

1.4. Outline. In section 2 we will state possible applications of Theorem 1.3 to various models with and without hard constraints. Along the way, we will give the necessary definitions and notations to remain self-contained. Thereafter, in section 3 we give the proofs of the main result Theorem 1.3 as well as all applications, i.e. Theorem 2.2, Corollary 2.3, Theorems 2.6 and 2.7.

2. APPLICATIONS

Our applications include two models of random graphs, namely the *exponential random graph models* and the *vertex-weighted exponential random graph models*, as well as models with hard constraints such as the *random coloring* or the *hard-core model*.

2.1. Exponential random graph models. In the last decades researchers have developed various models to describe real-world networks. Starting from the famous Erdős-Renyi model, which samples the presence or absence of edges independently, more sophisticated models have been proposed to explain certain observations which are not present in the Erdős-Renyi model, such as reciprocity in social networks, or local clustering, and hence incorporating a certain dependence structure. Among these are the exponential random graph models, which use ideas from statistical mechanics, very similar in spirit to Ising models. For a more thorough historical overview we refer to [BBS11] or the well-written survey [Cha16].

However, only recent works of S. Bhamidi, G. Bresler, A. Sly [BBS11] and S. Chatterjee and P. Diaconis [CD13] made progress in analyzing the Glauber dynamics associated to these models, as well as establishing large deviation principles. One of the main results is that in certain regimes of the parameter space (called the high temperature phase) the Glauber dynamics is rapidly mixing, whereas in the other regime (the low temperature phase) the Glauber dynamic takes exponential time

to reach equilibrium. However, the arguments in [BBS11] require the system to be monotone, i.e. the parameters to be positive.

We complement this by proving a (modified) logarithmic Sobolev inequality for the Glauber dynamics for a subset of the parameter space and as a consequence establish rapid mixing of the (continuous-time) Glauber dynamics. The method suggests that models with negative parameters should not behave differently from their monotone counterparts (where the parameter vector β is exchanged by its absolute value $|\beta|$).

The exponential random graph models are spin systems, parametrized by specifying certain graphs G_1, \dots, G_s and a distribution on the space of all graphs on n vertices (denoted by \mathcal{G}_n) by using the number of injections of the G_i as sufficient statistics. An easy example is given by taking G_1 to be the complete graph on 2 vertices and G_2 to be the complete graph on 3 vertices, and to sample a graph X on n vertices with probability $Z^{-1} \exp\left(\beta_1 E(X) + \frac{\beta_2}{n} T(X)\right)$, where $E(X)$ denotes the number of edges and $T(X)$ the number of triangles in the graph X .

More generally, for any two graphs G, H write $\mathcal{I}_G(H)$ for the set of graph homomorphism from G to H , i.e. all maps $\varphi : V(G) \rightarrow V(H)$ such that $v_i \sim_G v_j \Rightarrow \varphi(v_i) \sim_H \varphi(v_j)$, and let $N_G(H) = |\mathcal{I}_G(H)|$ be its cardinality; the normalized term $t(G, H) := \frac{N_G(H)}{|V(H)|^{|V(G)|}}$ is called the homomorphism density, and can be interpreted as the probability of a random mapping $\varphi : V(G) \rightarrow V(H)$ being a graph homomorphism. Let \mathcal{G}_n be the set of all graphs on n vertices, which we also identify with $\{0, 1\}^{n(n-1)/2}$.

Definition 2.1. Let $n \in \mathbb{N}, \beta = (\beta_1, \dots, \beta_s) \in \mathbb{R}^s$ and G_1, \dots, G_s be arbitrary, connected simple graphs with vertex set V_i and edge set E_i . The function

$$(2.1) \quad H_\beta(X) \equiv H(X) := n^2 \sum_{i=1}^s \beta_i \frac{N_{G_i}(X)}{n^{|V_i|}} = n^2 \sum_{i=1}^s \beta_i t(G_i, X),$$

is called Hamiltonian and the probability measure

$$(2.2) \quad \mu_\beta(\{X\}) = Z^{-1} \exp(H(X)) \quad \text{where} \quad Z = \sum_{X \in \mathcal{G}_n} \exp(H(X))$$

the exponential random graph model (ERGM) with parameters (β, G_1, \dots, G_s) , abbreviated as $\text{ERGM}(\beta, G_1, \dots, G_s)$.

It is customary to take $G_1 = K_2$ to be the complete graph on 2 vertices. For positive parameters β_i , the exponential random graph models assigns higher probability to graphs which contain G_i more often, whereas for negative β_i it favors the absence of G_i . For example, choosing the triangle as the only graph will result in graphs with lots of triangles, or more bipartite graphs, see e.g. [CD13, Figure 4].

To ease notation, we will not write the dependence of μ on G_1, \dots, G_s , since the graphs will be fixed. Moreover, we write for any vector $\beta = (\beta_1, \dots, \beta_s)$ its absolute value $|\beta|$ given by taking the absolute value of each component. Accordingly, $\mu_{|\beta|} = \text{ERGM}(|\beta|, G_1, \dots, G_s)$ is the associated monotone system. To avoid technicalities, we always assume that $n \geq \min_{i=2, \dots, s} |V_i|$, since otherwise we have $N_{G_i}(X) = 0$ for all $X \in \mathcal{G}_n$ and $i = 2, \dots, s$, in which case $\text{ERGM}(\beta, G_1, \dots, G_s)$ degenerates to an Erdős-Renyi random graph with parameters n and $\frac{e^{2\beta_1}}{1+e^{2\beta_1}}$.

Lastly, as is usual in the context of ERGM, for any set of parameters (β, G_1, \dots, G_s) we define the functions $\Phi_\beta, \varphi_\beta : [0, 1] \rightarrow \mathbb{R}$

$$(2.3) \quad \Phi_\beta(x) = \sum_{i=1}^s \beta_i |E_i| x^{|E_i|-1} = \beta_1 + \sum_{i=2}^s \beta_2 |E_i| x^{|E_i|-1}$$

$$(2.4) \quad \varphi_\beta(x) = \frac{\exp(2\Phi_\beta(x))}{1 + \exp(2\Phi_\beta(x))} = \frac{1}{2}(1 + \tanh(\Phi_\beta(x))).$$

We are now ready to prove the main theorem for exponential random graph models; recall the definition of the Dirichlet form (1.5).

Theorem 2.2. *Let μ_β be an ERGM(β, G_1, \dots, G_s) such that $\frac{1}{2}\Phi'_{|\beta|}(1) < 1$. The Glauber dynamics for μ_β satisfies*

$$(2.5) \quad \text{Ent}_{\mu_\beta}(e^f) \leq \frac{1}{2}C(\beta)n^2\mathcal{E}(e^f, f).$$

and is rapidly mixing.

Let us remark on these results.

Firstly, we are sure that the condition $\frac{1}{2}\Phi'_{|\beta|}(1) < 1$ is not optimal, since rapid mixing was proven for all exponential random graph models with positive parameters β_i and only one solution a^* to the equation $\varphi_\beta(a) = a$ satisfying $\varphi'(a^*) < 1$ (which the authors called *high temperature phase*, see [BBS11, Theorem 5]). Clearly $\frac{1}{2}\Phi'_{|\beta|}(1) < 1$ implies the uniqueness of a fixed point for $\varphi_{|\beta|}$ with $\varphi'(a) < 1$, so that the condition is strictly stronger than in [BBS11], but the result implies exponential decrease of the relative entropy and also applies to negative parameters. In [BBS11], the authors used a so-called burn-in phase for the Glauber dynamics due to the failure of path coupling in the case $\sup_{p \in [0,1]} \varphi'(p) > 1$, which we avoid by our requirements. Furthermore, note that the assumption $\Phi'_{|\beta|}(1) < 2$ is also present in [CD13, Theorem 6.2], where the authors show convergence in the cut-metric in probability to a mixture of Erdős-Renyi graphs in this region.

Secondly, with a slight modification of the proof one can show that under the assumptions of Theorem 2.2, a logarithmic Sobolev inequality holds with a slightly worse constant, which is however still of order n^2 . This is known to imply more properties than just rapid mixing, such as concentration of measure in the exponential random graph models. It remains an interesting open question whether a (modified) logarithmic Sobolev inequality with a constant of order n^2 holds in the full high temperature phase.

As an easy corollary we obtain sufficient conditions for exponential random graph models with two graphs $G_1 = K_2, G_2$ to be rapidly mixing.

Corollary 2.3. *Let G_2 be any connected simple graph with e_2 edges and assume $|\beta_2| < \frac{2}{e_2(e_2-1)}$. The Glauber dynamics of $\mu_\beta = \text{ERGM}(\beta_1, \beta_2, G_1, G_2)$ is rapidly mixing.*

Applying this to the star graph with k leaves S_k we obtain the sufficient condition $|\beta_2| < \frac{2}{k(k-1)}$, and for the triangle graph with $e_2 = 3$ this translates into $|\beta_2| < 1/3$.

By a more careful estimate, we can also prove the following Proposition.

Proposition 2.4. Let $\mu_\beta = \text{ERGM}(\beta, G_1, \dots, G_s)$, assume that $a^* \in [0, 1]$ satisfies $a^* = \varphi_{|\beta|}(a^*)$ and for $A^* = \max(a^*, 1 - a^*)$ we have

$$\gamma := \frac{1}{2} \left(\Phi'_{|\beta|}(a^*) + A^* \Phi''_{|\beta|}(1) \right) \left(\tanh'(\Phi_\beta(a^*)) + C_2 A^* \Phi'_{|\beta|}(1) \right) < 1.$$

Then the Glauber dynamics of μ_β is rapidly mixing.

Remark. The condition resembles the condition in [RR17, Theorem 1.5], expect for the fact that we have no $o(1)$ term which stems from the second order approximation of the tanh.

Corollary 2.5. Let $\mu_\beta = \text{ERGM}(\beta, G_1, \dots, G_s)$. If $a^* \in [0, 1]$ satisfies $a^* = \varphi_{|\beta|}(a^*)$ and

$$(2.6) \quad \gamma := \frac{1}{2} \left(\Phi'_{|\beta|}(a^*) + \Phi''_{|\beta|}(1) \right) \left(\tanh'(\Phi_\beta(a^*)) + C_2 \Phi_{(0, |\beta_2|, \dots, |\beta_s|)}(1) \right) < 1,$$

then the Glauber dynamics is rapidly mixing.

Remark. If we have the classical situation of a monotone system (see e.g. [BBS11; CD13; RR17]) that $\beta_2, \dots, \beta_s > 0$, we obtain the characterization

$$\|A\|_{1 \rightarrow 1} \leq \varphi'(a^*) + \frac{1}{2} \left(C_2 A^* \Phi'(a^*) \Phi'(1) + A^* \Phi''(1) (\tanh'(\Phi(a^*)) + C_2 A^* \Phi'(1)) \right)$$

and thus it is necessary for the Dobrushin uniqueness condition to have $\varphi'(a^*) < 1$, but with additional corrections due to the method.

2.2. Vertex-weighted exponential random graph models. Additionally, we are able to treat special cases of the vertex-weighted exponential random graph models as described in [DEY17]. The parameter-space is three-dimensional, i.e. $\beta = (\beta_1, \beta_2, p)$, and the model is given by the spin system on $\mathcal{Y} = \{0, 1\}^n$ via the Hamiltonian

$$H(\sigma) := \log \left(\frac{p}{1-p} \right) \sum_i \sigma_i + \frac{\beta_1}{n} \sum_{i \neq j} \sigma_i \sigma_j + \frac{\beta_2}{n^2} \sum_{i \neq j \neq k} \sigma_i \sigma_j \sigma_k,$$

which resembles the Hamiltonian in the exponential random graph model. We define the function

$$\varphi_\beta(\lambda) := \frac{\exp(h_\beta(\lambda))}{1 + \exp_\beta(h(\lambda))} = \frac{\exp(\beta_1 \lambda + \beta_2 \lambda^2 + \log(p/(1-p)))}{1 + \exp(\beta_1 \lambda + \beta_2 \lambda^2 + \log(p/(1-p)))}.$$

Theorem 2.6. If the parameter $\beta := (\beta_1, \beta_2, p)$ satisfies

$$(2.7) \quad \sup_{\lambda \in (0,1)} |\varphi'_\beta(\lambda)| < 1,$$

then a modified logarithmic Sobolev inequality holds and the Glauber dynamics is rapidly mixing.

2.3. Random coloring model. The graph models considered thus far are spin systems μ with no hard constraints, i.e. any configuration is *admissible* (has positive probability). Certain models, however, are supported on a strict subset $\Omega_0 \subset \mathcal{X}^{\mathcal{I}}$.

To obtain mixing time estimates for models with hard constraints, we shall pursue a two-step strategy. Firstly, we change the probability space from Ω_0 to $\mathcal{Y} = \mathcal{X}^{\mathcal{I}}$ by setting $\mu(x) = 0$ for all $x \in \mathcal{Y} \setminus \Omega_0$ to apply Theorem 1.2, and estimate the right hand side of equation (1.8) for the choice $f = e^g$ as in the proof of Theorem 1.3. In the second step, we restrict again to functions $f : \Omega_0 \rightarrow \mathbb{R}_+$ (since both sides

on the inequality only depend on $x \in \text{supp } \mu$) and identify the right hand side as the Dirichlet form associated to the Glauber dynamics on Ω_0 , hence establishing a modified logarithmic Sobolev inequality, from which we infer the mixing times estimates.

To this end, we briefly introduce the random k -coloring model. Given a finite graph $G = (V, E)$ with maximum degree Δ and a finite set of colors $C = \{1, \dots, k\}$, the configuration space in this model is the set of all proper colorings $\Omega_0 \subset C^V$, i.e. the set of all $\varphi \in C^V$ such that $v \sim w \Rightarrow \varphi_v \neq \varphi_w$, and $\mu = \mu(G, C)$ denotes the uniform distribution on Ω_0 .

The Glauber dynamics for a sequence of bounded-degree graphs was shown to be rapidly mixing by M. Jerrum [Jer95] for $k \geq 2\Delta + 1$ via a path coupling approach. We recover these results using the entropy approach. Again, we consider the (continuous-time) Glauber dynamics with respect to μ .

Theorem 2.7. *Let $G_n = (V_n, E_n)$ be a sequence of graphs with uniformly bounded maximum degree Δ and $k \geq 2\Delta + 1$ be fixed. The (continuous-time) Glauber dynamics $(Y_t)_{t \geq 0}$ on Ω_0 is rapidly mixing.*

2.4. Hard-core model. Another model with hard constraints is the hard-core model with fugacity λ . Given a graph $G = (V, E)$ with maximum degree Δ , the hard-core model is the spin system on $\mathcal{Y} = \{0, 1\}^V$ which assigns probability $Z^{-1} \lambda^{|\sigma|}$ to any admissible configuration, i.e. any configuration such that $\sigma_v \sigma_w = 0$ for all $v \sim w$. The parameter λ is called *fugacity*. It was shown in [Vig01, Theorem 1] that if $G_n = (V_n, E_n)$ is a sequence of graphs with uniformly bounded degree Δ and $\lambda < \frac{2}{\Delta-2}$, then the Glauber dynamics is rapidly mixing. We can recover a partial result.

Theorem 2.8. *Let $G_n = (V_n, E_n)$ be a sequence of graphs with bounded maximum degree Δ and let $\lambda < \frac{1}{\Delta-1}$. The Glauber dynamics corresponding to $\mu_{G_n, \lambda}$ is rapidly mixing.*

Interestingly, with methods closer to the Bakry-Emery theory and a characterization of Ricci curvature for Markov chains as developed by J. Maas [Maa11] and A. Mielke [Mie13], M. Erbar, C. Henderson, G. Menz and P. Tetali [Erb+17] have shown for the hard-core model a positive Ricci curvature under the assumption $\lambda \leq \frac{1}{\Delta}$, which also implies a modified logarithmic Sobolev inequality.

3. PROOFS

In this section we will prove our main result, Theorem 1.3, and apply it to the exponential random graph model to prove Theorem 2.2, the vertex-weighted ERGM to prove Theorem 2.6, the random coloring model to prove Theorem 2.7 and lastly the hard-core model to prove Theorem 2.8.

3.1. Proofs of main results.

Proof of Theorem 1.3. Let us define $\Omega_0 := \text{supp } \mu$, where supp is the support of μ , i.e. $\text{supp}(\mu) := \{y \in \mathcal{Y} : \mu(y) > 0\}$. We can apply Theorem 1.2 to obtain for any $f : \mathcal{Y} \rightarrow \mathbb{R}$ vanishing outside of Ω_0

$$(3.1) \quad \text{Ent}_\mu(f) \leq \frac{2}{\alpha_1 \alpha_2} \sum_{i \in \mathcal{I}} \int \text{Ent}_{\mu(\cdot | \bar{x}_i)}(f(\bar{x}_i, \cdot)) d\mu(x).$$

This is equivalent to the fact that on the probability space (Ω_0, μ) , any function $f : \Omega_0 \rightarrow \mathbb{R}_+$ satisfies the same inequality, which we shall work with from now on. For any probability measure $(\Omega, \mathcal{F}, \nu)$ and any function f such that $f, e^f \in L^2(\nu)$, we have by Jensen's inequality and the symmetry in the covariance

$$(3.2) \quad \text{Ent}_\nu(e^f) \leq \text{Cov}_\nu(f, e^f) = \int \left(\int (f(y) - f(x)) d\nu(x) \right) e^{f(y)} d\nu(y).$$

Apply the inequality (3.2) in the integral on the right hand side of equation (3.1) to get

$$(3.3) \quad \text{Ent}_\mu(e^f) \leq \frac{2}{\alpha_1 \alpha_2^2} \sum_{i \in \mathcal{I}} \int \left(\int (f(x) - f(\bar{x}_i, y)) d\mu(y | \bar{x}_i) \right) e^{f(x)} d\mu(x).$$

Finally, observe that for the transition matrix P and the generator $-L = I - P$ of the Glauber dynamics (on Ω_0) we have

$$\begin{aligned} \mathcal{E}(e^f, f) &= \mathbb{E}_\mu(e^f(-Lf)) = \int \sum_{y \in \Omega_0} (f(x) - f(y)) P(x, y) e^{f(x)} d\mu(x) \\ &= \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} \iint (f(x) - f(\bar{x}_i, y)) d\mu(y | \bar{x}_i) e^{f(x)} d\mu(x), \end{aligned}$$

so that a normalization of inequality (3.3) by $|\mathcal{I}|$ leads to

$$(3.4) \quad \text{Ent}_\mu(e^f) \leq 2 \frac{|\mathcal{I}|}{\alpha_1 \alpha_2^2} \mathcal{E}(e^f, f),$$

and the modified logarithmic Sobolev inequality is established.

Now let $(\mu_n)_n$ be a sequence of spin systems with sites $(\mathcal{I}_n)_n$, spins \mathcal{X} , and define $\mathcal{Y}_n = \text{supp}(\mu_n) \subset \mathcal{X}^{\mathcal{I}_n}$. To prove rapid mixing, note that

$$\frac{2}{\rho_0} = \inf \left\{ \frac{\mathcal{E}(e^f, f)}{\text{Ent}_{\mu_n}(e^f)} : f \neq \text{const} \right\} \geq \frac{\alpha_1 \alpha_2^2}{2|\mathcal{I}_n|}.$$

If we denote $\mu_n^* = \min_{y \in \mathcal{Y}_n} \mu_n(y)$, by Theorem 1.1 this leads to

$$d_{TV}(\delta_y * P^t, \mu_n)^2 \leq 2 \log(1/\mu_n^*) \exp(-2\rho_0^{-1}t) \leq 2 \log(1/\mu_n^*) \exp(-\alpha_1 \alpha_2^2 (2|\mathcal{I}_n|)^{-1}t).$$

Hence for $t \geq \frac{2|\mathcal{I}_n|}{\alpha_1 \alpha_2^2} \cdot (\log 2 + 2 + \log \log(1/\mu_n^*))$ we have for any $y \in \mathcal{Y}_n$

$$d_{TV}(\delta_y * P^t, \mu_n)^2 \leq e^{-2},$$

i.e. $t_{\text{mix}}(n) \leq \frac{2|\mathcal{I}_n|}{\alpha_1 \alpha_2^2} \cdot (\log 2 + 2 + \log \log(1/\mu_n^*))$. Consequently, to establish rapid mixing it remains to show $\log \log 1/\mu_n^* = O(\log |\mathcal{I}_n|)$, but this is easy using the definition of α_1 , since by conditioning and iterating we obtain for any $y \in \mathcal{Y}_n$

$$\frac{1}{\mu_n(y)} = \frac{1}{\mu_n(y_i | \bar{y}_i)} \mu_n(\bar{y}_i)^{-1} \leq \alpha_1^{-1} \mu(\bar{y}_i)^{-1} \leq \alpha_1^{-|\mathcal{I}_n|}.$$

□

Proof of Corollary 1.4. We have already shown that $I(\mu_n) = \tilde{\beta}(\mu_n)$ for a measure μ_n with full support. Hence this is simply a rephrasing of the conditions. □

3.2. Proofs of the applications. It is convenient to introduce some notation for the exponential random graph models. Let \mathcal{G}_n denote the set of all graphs on n vertices and for any $X \in \mathcal{G}_n$ and any edge $e = (i, j) \in \mathcal{I}_n := \{(i, j) \in \{1, \dots, n\}^2 : i < j\}$ let X_{e+} be the graph with edge set $E(X_{e+}) = E(X) \cup e$ and X_{e-} with edge set $E(X_{e-}) = E(X) \setminus e$. For any function $f : \mathcal{G}_n \rightarrow \mathbb{R}$ we define the discrete derivative in the e -th direction as

$$\mathfrak{d}_e f(X) = f(X_{e+}) - f(X_{e-}).$$

More generally, given edges e_1, \dots, e_k we define $\mathfrak{d}_{e_1 \dots e_k}$ recursively, i.e. $\mathfrak{d}_{e_1 \dots e_k} f = \mathfrak{d}_{e_1}(\mathfrak{d}_{e_2 \dots e_k} f)$. It is easy to see that the definition does not depend on the order of the edges and $\mathfrak{d}_{ee} f = 0$. The partial derivatives of the Hamiltonian are given by

$$\mathfrak{d}_e H(X) = 2\beta_1 + n^2 \sum_{i=2}^s \frac{\beta_i}{n^{|V_i|}} (N_{G_i}(X_{e+}) - N_{G_i}(X_{e-}))$$

Now we use the fact if G_i injects into X_{e-} , then it also injects into X_{e+} , and hence the sum is only nonzero if the edge e is essential for the injection, and write $N_{G_i}(X, e)$ to denote the number of injections of G_i into X which use the edge $e \in E(X)$, so that $\mathfrak{d}_e H(X) = 2\beta_1 + n^2 \sum_{i=2}^s \frac{\beta_i}{n^{|V_i|}} N_{G_i}(X, e)$. Especially it can be easily seen that $|\mathfrak{d}_e H(X)| = O(1)$.

Proof of Theorem 2.2. We want to apply Theorem 1.3 in the form of Corollary 1.4. The spin system is given by $\mathcal{Y}_n := \{0, 1\}^{\mathcal{I}_n}$, and μ_n is the push-forward of the measure associated to the exponential random graph model $\text{ERGM}(\beta, G_1, \dots, G_s)$ on \mathcal{G}_n . The condition of (1.13) is easy to check, since for any $e \in \mathcal{I}_n$ and any $y \in \mathcal{Y}_n$

$$\mu_n(y_e \mid \bar{y}_e) = \frac{1}{2} (1 + \tanh(\mathfrak{d}_e H(y)/2))$$

and $\mathfrak{d}_e H(y) = O(1)$, where the constant depends on $(|\beta|, G_1, \dots, G_s)$ only. Hence it remains to prove the condition (1.14). To this end, let again $x = x_{f+}, y = x_{f-}$ be two graphs which differ in one edge f only, and observe that for each other edge e

$$\begin{aligned} d_{TV}(\mu_n(\cdot \mid \bar{x}_e), \mu_n(\cdot \mid \bar{y}_e)) &= |\mu_n(1 \mid \bar{x}_e) - \mu_n(1 \mid \bar{y}_e)| \\ &= \frac{1}{2} |\tanh(\mathfrak{d}_e H(x_{f+})/2) - \tanh(\mathfrak{d}_e H(x_{f-})/2)| \\ &\leq \frac{1}{4} |\mathfrak{d}_{fe} H(x)| \leq \frac{n^2}{4} \sum_{i=2}^s |\beta_i| \frac{N_{G_i}(x, f, e)}{n^{|V_i|}} \\ &\leq \frac{n^2}{4} \sum_{i=2}^s |\beta_i| \frac{N_{G_i}(K_n, f, e)}{n^{|V_i|}}, \end{aligned}$$

i.e. $J_{fe} \leq \frac{n^2}{4} \sum_{i=2}^s |\beta_i| \frac{N_{G_i}(K_n, f, e)}{n^{|V_i|}}$. Thus after summation in $f \in \mathcal{I}_n$ we obtain by [BBS11, Lemma 9(c)]

$$\sum_{f \neq e} J_{fe} \leq \frac{n^2}{4} \sum_{i=2}^s |\beta_i| \sum_{f \neq e} \frac{N_{G_i}(K_n, f, e)}{n^{|V_i|}} = \frac{n^2}{4} \sum_{i=2}^s |\beta_i| \frac{2|E_i|(|E_i| - 1)n^{|V_i|-2}}{n^{|V_i|}} = \frac{1}{2} \Phi'_{|\beta|}(1).$$

Since the right-hand side is independent of $e \in \mathcal{I}_n$, this immediately yields

$$\|J\|_{1 \rightarrow 1} \leq \frac{1}{2} \Phi'_{|\beta|}(1) < 1.$$

Moreover, since J is a symmetric matrix, we have $\|J\|_{2 \rightarrow 2} \leq \|J\|_{1 \rightarrow 1}$, showing the modified logarithmic Sobolev inequality and the rapid mixing. \square

Proof of Corollary 2.3. The proof is trivial, since $\frac{1}{2}\Phi'_{|\beta|}(1) = \frac{1}{2}|\beta_2|e_2(e_2 - 1) < 1$, and thus the Corollary follows from Theorem 2.2. \square

Proof of Proposition 2.4. As in the proof of Theorem 2.2 it remains to check Dobrushin's uniqueness condition (1.10) for the measure μ_β . The proof is a modification of the proof of [RR17, Lemma 3.1], however we will only use a first order expansion of the tanh function instead of a second-order expansion.

Fix two edges $f = (m, n)$ and $e = (k, l)$ and two graphs X, Y which differ only in f . Using the Taylor approximation for some $s \in (0, 1)$

$$\begin{aligned} d_{TV}(\mu_\beta(\cdot | \overline{X}_e), \mu_\beta(\cdot | \overline{Y}_e)) &= \frac{1}{2} \left| \tanh\left(\frac{1}{2}\mathfrak{d}_e H(X_{f+})\right) - \tanh\left(\frac{1}{2}\mathfrak{d}_e H(X_{f-})\right) \right| \\ &\leq \frac{1}{4} |\mathfrak{d}_{fe} H(X)| \tanh'\left(\frac{s}{2}\mathfrak{d}_e H(X_{f+}) + \frac{1-s}{2}\mathfrak{d}_e H(X_{f-})\right) \\ &=: \frac{1}{4} I_1(f, e) \cdot I_2. \end{aligned}$$

We will bound $I_1(f, e)$ and I_2 separately. To bound I_2 , by adding and subtracting $\tanh'(s(\Phi(a^*) + (1-s)\Phi(a^*)))$ and using $|\tanh'(a) - \tanh'(b)| \leq C_2|a - b|$ we get

$$I_2 \leq \tanh'(\Phi(a^*)) + sC_2|\mathfrak{d}_e H(X_{f+})/2 - \Phi(a^*)| + (1-s)C_2|\mathfrak{d}_e H(X_{f-})/2 - \Phi(a^*)|,$$

and since

$$(3.5) \quad |\mathfrak{d}_e H(X_{f\pm})/2 - \Phi(a^*)| \leq \sum_{i=2}^s |\beta_i| |E_i| (|E_i| - 1) A^* = A^* \Phi'_{|\beta|}(1)$$

we obtain

$$(3.6) \quad I_2 \leq \tanh'(\Phi_\beta(a^*)) + C_2 A^* \Phi'_{|\beta|}(1).$$

As for I_1 , we make use of the last part of the proof of [RR17, Lemma 3.1] to get

$$(3.7) \quad \frac{1}{2} \sum_{f \neq e} I_1(e, f) \leq \Phi'_{|\beta|}(a^*) + A^* \Phi''_{|\beta|}(1).$$

Thus, combining (3.6) and (3.7) leads to

$$(3.8) \quad \|A\|_{1 \rightarrow 1} \leq \frac{1}{2} (\Phi'_{|\beta|}(a^*) + A^* \Phi''_{|\beta|}(1)) (\tanh'(\Phi_\beta(a^*)) + C_2 A^* \Phi'_{|\beta|}(1)).$$

Again, by symmetry this implies $\|A\|_{2 \rightarrow 2} < 1$, and so the result follows from Corollary 1.4. \square

Remark. In the proof of Proposition 2.4 it will be clear that the estimate (3.5) can be improved for certain ERGM(β, G_1, \dots, G_s). Indeed, since $\mathfrak{d}_e N_{G_i}(X)$ is the number of injections of G_i into X using the edge e , it can be easily shown that $0 \leq n^2 \frac{\mathfrak{d}_e N_{G_i}(X)}{2e_i n^{|V_i|}} \leq 1$ (and $0 \leq (a^*)^{e_i-1} \leq 1$), so

$$(3.9) \quad \left| \frac{\mathfrak{d}_e H(X_{f+})}{2} - \sum_{i=1}^s \beta_i e_i (a^*)^{e_i-1} \right| \leq \sum_{i=2}^s |\beta_i| e_i \left| \frac{\mathfrak{d}_e N_{G_i}(X)}{2e_i (n)^{|V_i|-2}} - (a^*)^{e_i-1} \right| \leq \sum_{i=2}^s |\beta_i| e_i.$$

Since $A^* \geq \frac{1}{2}$, for any ensemble of graphs of which H_2 (the 2-star) is not a part, this is superior as

$$A^* \Phi'_{|\beta|}(1) = A^* \sum_{i=2}^s |\beta_i| |E_i| (|E_i| - 1) \geq 2A^* \sum_{i=2}^s |\beta_i| |E_i| \geq \sum_{i=2}^s |\beta_i| |E_i|.$$

For the best bound, one can use a combination thereof, bounding β_2 (corresponding to H_2) via A^* and the rest as above. However, since calculating A^* requires solving the equation $\tanh(P(a)) = 2a - 1$ for a polynomial P of degree $\max_i e_i$, this is usually intractable, and thus one uses the inequality $A^* \leq 1$, and equation (3.9) is better.

Moreover the estimate (3.8) can sometimes also be improved by ignoring the estimates for I_2 in the proof and simply using $\tanh'(x) \leq 1$, leading to

$$(3.10) \quad \|A\|_{1 \rightarrow 1} \leq \frac{1}{2} \left(\Phi'_{|\beta|}(a^*) + A^* \Phi''_{|\beta|}(1) \right)$$

Next, let us prove the statement for vertex-weighted exponential random graph models.

Proof of Theorem 2.6. First note that we have for fixed parameter $\beta = (\beta_1, \beta_2, p)$

$$\mu(x) := \mu_\beta(x) := Z^{-1} \exp \left(\frac{\beta_1}{n} \sum_{i \neq j} x_i x_j + \frac{\beta_2}{n^2} \sum_{i \neq j \neq k} x_i x_j x_k + \log \frac{p}{1-p} \sum_{i=1}^n x_i \right).$$

Let us define $H_n(x) := \frac{\beta_1}{n} \sum_{i \neq j} x_i x_j + \frac{\beta_2}{n^2} \sum_{i \neq j \neq k} x_i x_j x_k + \log \frac{p}{1-p} \sum_{i=1}^n x_i$. Moreover, since $x_i \in \{0, 1\}$ implies $x_i^k = x_i$ for all $k \in \mathbb{N}$, we can rewrite this using $S := \sum_{i=1}^n x_i$ as

$$\mu(x) = Z^{-1} \exp \left(\frac{\beta_1}{n} S(S-1) + \frac{\beta_2}{n^2} S(S-1)(S-2) + \log \frac{p}{1-p} S \right).$$

Hence for $\mathcal{X} := \{0, 1\}$ and $\mathcal{I}_n := \{1, \dots, n\}$ we are in the situation of Theorem 1.3, and it remains to check conditions (1.9) and (1.10). Observe that we have (with the same notations as in the exponential random graph models)

$$\mu(1 \mid \bar{x}_e) = \frac{\exp(\mathfrak{d}_e H_n(\bar{x}_e, 1))}{1 + \exp(\mathfrak{d}_e H_n(\bar{x}_e, 1))} = \frac{1}{2} (1 + \tanh(\mathfrak{d}_e H_n(x)/2)),$$

where in this case $|\mathfrak{d}_e H_n(x)| = |\frac{2\beta_1}{n} \sum_{i \neq e} x_i + \frac{3\beta_2}{n^2} \sum_{i \neq j, i, j \neq e} x_i x_j + \log(p/(1-p))|$ is bounded by a constant depending on β , so that a lower bound on the conditional probabilities holds. The inequality (1.10) is already implicitly proven in the proof of [DEY17, Lemma 6], which we modify. Fix a site $e \in \mathcal{I}_n$ and two configurations x, y differing solely at $f \in \mathcal{I}_n$, i.e. $x_f = 1, y_f = 0$, and let $S := \sum_{i=1}^n y_i$. We have

$$d_{TV}(\mu(\cdot \mid \bar{x}_e), \mu(\cdot \mid \bar{y}_e)) = \frac{1}{2} |\tanh(\mathfrak{d}_e H_n(\bar{x}_e, 1)) - \tanh(\mathfrak{d}_e H_n(\bar{y}_e, 1))|$$

and since H_n (and as a consequence $\mathfrak{d}_e H_n$) only depends on the sum S of a vector, by defining $h_n(\lambda) := \beta_1 \lambda + \beta_2 \lambda^2 - \frac{\beta_2}{n} \lambda + \log(p/(1-p))$ we can estimate for some $\xi \in (0, 1)$

$$(3.11) \quad J_{fe} \leq \left| \frac{\exp(h_n((S+1)/n))}{1 + \exp(h_n((S+1)/n))} - \frac{\exp(h_n(S/n))}{1 + \exp(h_n(S/n))} \right| = \frac{1}{n} \left| \left(\frac{\exp \circ h_n}{1 + \exp \circ h_n} \right)'(\xi) \right|.$$

Lastly, if we define $h(\lambda) = \beta_1 \lambda + \beta_2 \lambda^2 + \log(p/(1-p))$, using the Lipschitz property of the function $\exp(x)/(1 + \exp(x))$ it can be shown that

$$\left| \frac{\exp \circ h_n}{1 + \exp \circ h_n} - \frac{\exp \circ h}{1 + \exp \circ h} \right| = O(n^{-1})$$

and h_n can be replaced by h in (3.11) with an error of $O(n^{-2})$. By summing up over $f \neq e$, we obtain for n large enough and all parameters such that

$$(3.12) \quad \sup_{\lambda \in (0,1)} \left| \frac{\exp \circ h}{1 + \exp \circ h} \right| < 1$$

that the inequality (1.10) holds. \square

Remark. Note that the condition (3.12) can be written in terms of the functions defined for exponential random graphs. More specifically, we have for any $x \in \mathbb{R}$

$$\frac{\exp(x)}{1 + \exp(x)} = \frac{1}{2}(1 + \tanh(x/2)),$$

and hence this functions corresponds to φ_β for an ERGM(β, K_1, K_2, K_3) with $\beta_1 = \log(p/(1-p)), \beta_2 = \frac{\alpha_1}{2}, \beta_3 = \frac{\alpha_2}{6}$.

Proof of Theorem 2.7. This Theorem is again an application of Theorem 1.3. Let us first show that

$$J_{v,w} := \frac{1}{\Delta + 1} \mathbb{1}_{v \sim w}$$

can be used as an interdependence matrix (regardless of $n \in \mathbb{N}$). To see this, let $c^1, c^2 \in \Omega_0$ be two colorings that differ only in one vertex v_1 , and v_2 be another vertex. In the case $v_1 \sim v_2$ (in G_n) the measures $\mu_{v_2}(\cdot | \overline{c_{v_2}^i})$ are uniform on $C \setminus \{c_{v_k}^i : v_k \sim v_1\}$ for $i = 1, 2$, and hence

$$\begin{aligned} d_{TV}(\mu_{v_2}(\cdot | \overline{c_{v_2}^1}), \mu_{v_2}(\cdot | \overline{c_{v_2}^2})) &= \frac{1}{2} \left(\frac{1}{k - |\{c_{v_2}^1 : v_2 \sim v_1\}|} + \frac{1}{k - |\{c_{v_2}^2 : v_2 \sim v_1\}|} \right) \\ &\leq \frac{1}{k - \Delta} \leq \frac{1}{\Delta + 1}. \end{aligned}$$

On the other hand, if $v_2 \not\sim v_1$, then $\mu_{v_2}(\cdot | \overline{c_{v_2}^i})$ are equal and thus $J_{v_1, v_2} = 0$.

Since J is a symmetric matrix, we obtain

$$\|J\|_{2 \rightarrow 2} \leq \|J\|_{1 \rightarrow 1} \leq \max_{v_j \in V_n} \sum_{v_i \in V_n} J_{v_i, v_j} \leq \frac{\Delta}{\Delta + 1} < 1.$$

Moreover, we have to show that $\tilde{\beta}(\mu_n) \geq \alpha_1$ uniformly in $n \in \mathbb{N}$. Let $S \subsetneq V_n, S \neq \emptyset$ be arbitrary, $v_1 \notin S$ and $c^S \in C^S$ be a proper coloring of $G|_S = (S, E_n \cap S \times S)$ and $c^{v_1} \in C \setminus \{c_{v_2} : v_2 \in S, v_2 \sim v_1\}$. Using the definition $\Omega_0(G)$ for the set of all proper colorings of an arbitrary graph G (with a fixed number of colors, here k), we have

$$(3.13) \quad \mu_{\overline{S}}(c^{v_1} | c^S) = \frac{\mu(c^{v_1}, c^S)}{\mu(c^S)} = \frac{|\Omega_0(G|_S)|}{|\Omega_0(G|_{S \cup v_1})|}.$$

It is clear that $|\Omega_0(G|_S)| = \frac{1}{|C|} |\Omega_0(\widetilde{G}_S)|$, where \widetilde{G}_S is obtained by adding an isolated vertex v_1 to S . Hence we fix the vertex set $S \cup v_1$ and rewrite equation (3.13) as follows. Let $N(v_1, S) = \{v_2 \in S : v_2 \sim v_1\} = \{e_1, \dots, e_l\}$ be the neighbors of v_1 in S and for any $e_1, \dots, e_k \in N(v_1, S)$ let G_{e_1, \dots, e_k} be the graph with edge set $(E_n \cap S \times S) \cup \{e_1, \dots, e_k\}$, so that

$$\mu_{\overline{S}}(c^{v_1} | c^S) = \frac{1}{|C|} \prod_{k=1}^l \frac{|\Omega_0(G_{e_1, \dots, e_{k-1}})|}{|\Omega_0(G_{e_1, \dots, e_k})|}.$$

By [Jer95, equation (2)], it follows that each of the ratios is bounded from below by a constant depending on Δ , thus resulting in

$$\mu_{\bar{S}}(c^{v_1} \mid c^S) \geq \frac{1}{|C|} c(\Delta),$$

with a possible choice $c(\Delta) = \left(\frac{\Delta+1}{\Delta+2}\right)^\Delta$. The case $S = \emptyset$ is easier, since $\mu_i(c^i) = \frac{1}{|C|}$ by the invariance of the random coloring model induced by a relabeling of the colors C . \square

Proof of Theorem 2.8. Since we are going to require hard-core models corresponding to various graphs, we will write μ_G to emphasize the graph under consideration. The fugacity λ will not change. Let us show that $J_{v_1, v_2} = \frac{\lambda}{1+\lambda} \mathbb{1}_{v_1 \sim v_2}$ can be used as an interdependence matrix. Let $v_1 \in V$ be a site, $\sigma, \sigma^2 \in \mathcal{Y}$ be two admissible configurations differing only at site v_1 (without loss of generality $\sigma_{v_1}^1 = 1, \sigma_{v_1}^2 = 0$), and $v_2 \in V$ be another site. If $v_2 \sim v_1$, then $\mu_G(1 \mid \sigma_{v_1}^1) = 0$, whereas $\mu_G(1 \mid \sigma_{v_1}^2) = \frac{\lambda}{1+\lambda}$. If $v_2 \not\sim v_1$ we have $\mu_G(\cdot \mid \sigma_{v_1}^1) = \mu_G(\cdot \mid \sigma_{v_1}^2)$. Hence by the symmetry of J

$$\|J\|_{2 \rightarrow 2} \leq \|J\|_{1 \rightarrow 1} \leq \Delta \frac{\lambda}{1+\lambda} < 1$$

since $\lambda < \frac{1}{\Delta-1}$.

To see that there is a lower bound on the conditional probabilities, let us first consider the case $S = \emptyset$. Let $v \in V$ be arbitrary, write $N(v)$ for the neighborhood of v and A for the complement of $v \cup N(v)$, and observe that

$$\begin{aligned} \mu_G(\sigma_v = 1) &= \mu(\sigma_v = 1, \sigma_{N(v)} = 0) = Z^{-1} \sum_{\tilde{\sigma}_A} \mu(\sigma_v = 1, \sigma_{N(v)} = 0, \sigma_A = \tilde{\sigma}_A) \\ &= \lambda Z^{-1} \sum_{\tilde{\sigma}_A \text{ adm}} \lambda^{|\tilde{\sigma}_A|} =: \lambda Z^{-1} Z_A. \end{aligned}$$

where $\tilde{\sigma}_A$ ranges over all admissible configurations. Note that due to $\sigma_{N(v)} = 0$ these are actually all admissible configurations of the graph $G|_A = (A, E \cap A \times A)$. The normalizing constant can be bounded from above and below by

$$Z = \sum_{\tilde{\sigma}_A \text{ adm.}} \lambda^{|\tilde{\sigma}_A|} \sum_{\tilde{\sigma}_{A^c}} \lambda^{|\tilde{\sigma}_{A^c}|} \mathbb{1}_{(\tilde{\sigma}_A, \tilde{\sigma}_{A^c}) \text{ adm.}} \leq 2^{\Delta+1} Z_A$$

and

$$Z \geq (\lambda + 1) Z_A,$$

which follows by only considering the configurations $\sigma_v = 1, \sigma_{N(v)} = 0$ and $\sigma_{v \cup N(v)} = 0$. As a consequence, we have

$$(3.14) \quad \frac{\lambda}{2^{\Delta+1}} \leq \mu(\sigma_i = 1) \leq \frac{\lambda}{\lambda + 1}.$$

The case $S \neq \emptyset$ follows by a reduction argument. Let $\tilde{\sigma}_S$ be an admissible configuration of $G|_S$ and let $T := \{w \in S : \sigma_w = 0\} \subset S$ be the free sites in S . By explicitly writing the conditional probability one can see that for any configuration σ_{S^c} and any $v \notin S$ we have

$$\mu_G(\sigma_v = 1 \mid \sigma_S = \tilde{\sigma}_S) = \mu_{G|_{V \setminus T}}(\sigma_v = 1 \mid \sigma_{S \setminus T} = (1, \dots, 1)).$$

Now the graph $G|_{V \setminus T}$ can be divided into three parts: $(T, N(T), R)$, where $N(T) = \cup_{v \in T} N(v)$ and

$$\mu_G(\sigma_v = 1 \mid \sigma_S = \tilde{\sigma}_S) = \mu_{G|_R}(\sigma_v = 1),$$

which has an upper and lower bound by inequality (3.14). Thus $\tilde{\beta}(\mu_G) \geq c(\Delta, \lambda)$ and the Theorem follows from Theorem 1.3. \square

REFERENCES

- [BBS11] Shankar Bhamidi, Guy Bresler, and Allan Sly. “Mixing time of exponential random graphs”. In: *Ann. Appl. Probab.* 21.6 (2011), pp. 2146–2170. DOI: 10.1214/10-AAP740.
- [BL98] S. G. Bobkov and M. Ledoux. “On modified logarithmic Sobolev inequalities for Bernoulli and Poisson measures”. In: *J. Funct. Anal.* 156.2 (1998), pp. 347–365. DOI: 10.1006/jfan.1997.3187.
- [BT06] Sergey G. Bobkov and Prasad Tetali. “Modified logarithmic Sobolev inequalities in discrete settings”. In: *J. Theoret. Probab.* 19.2 (2006), pp. 289–336. DOI: 10.1007/s10959-006-0016-3.
- [BLM13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities. A nonasymptotic theory of independence*, With a foreword by Michel Ledoux. Oxford University Press, Oxford, 2013, pp. x+481. ISBN: 978-0-19-953525-5.
- [Cha16] Sourav Chatterjee. “An introduction to large deviations for random graphs”. In: *Bull. Amer. Math. Soc. (N.S.)* 53.4 (2016), pp. 617–642. DOI: 10.1090/bull/1539.
- [CD13] Sourav Chatterjee and Persi Diaconis. “Estimating and understanding exponential random graph models”. In: *Ann. Statist.* 41.5 (2013), pp. 2428–2461. DOI: 10.1214/13-AOS1155.
- [DEY17] Ryan DeMuse, Terry Easlick, and Mei Yin. “Mixing Time of Vertex-Weighted Exponential Random Graphs”. In: *arXiv preprint* (2017). arXiv: 1712.07568.
- [DS96] Persi Diaconis and Laurent Saloff-Coste. “Logarithmic Sobolev inequalities for finite Markov chains”. In: *Ann. Appl. Probab.* 6.3 (1996), pp. 695–750. DOI: 10.1214/aop/1034968224.
- [Erb+17] Matthias Erbar, Christopher Henderson, Georg Menz, and Prasad Tetali. “Ricci curvature bounds for weakly interacting Markov chains”. In: *Electron. J. Probab.* 22 (2017), Paper No. 40, 23. DOI: 10.1214/17-EJP49.
- [GQ03] Fuqing Gao and Jeremy Quastel. “Exponential decay of entropy in the random transposition and Bernoulli-Laplace models”. In: *Ann. Appl. Probab.* 13.4 (2003), pp. 1591–1600. DOI: 10.1214/aop/1069786512.
- [GSS18] Friedrich Götze, Holger Sambale, and Arthur Sinulis. “Higher order concentration for functions of weakly dependent random variables”. In: *arXiv preprint* (2018). arXiv: 1801.06348.
- [Gro75] Leonard Gross. “Logarithmic Sobolev inequalities”. In: *Amer. J. Math.* 97.4 (1975), pp. 1061–1083. DOI: 10.2307/2373688.
- [Jer95] Mark Jerrum. “A very simple algorithm for estimating the number of k -colorings of a low-degree graph”. In: *Random Structures Algorithms* 7.2 (1995), pp. 157–165. DOI: 10.1002/rsa.3240070205.

- [Led01] Michel Ledoux. *The concentration of measure phenomenon*. Vol. 89. Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, 2001, pp. x+181. ISBN: 0-8218-2864-9.
- [Maa11] Jan Maas. “Gradient flows of the entropy for finite Markov chains”. In: *J. Funct. Anal.* 261.8 (2011), pp. 2250–2292. DOI: 10.1016/j.jfa.2011.06.009.
- [Mar15] Katalin Marton. “Logarithmic Sobolev inequalities in discrete product spaces: a proof by a transportation cost distance”. In: *arXiv preprint* (2015). arXiv: 1507.02803.
- [Mie13] Alexander Mielke. “Geodesic convexity of the relative entropy in reversible Markov chains”. In: *Calc. Var. Partial Differential Equations* 48.1-2 (2013), pp. 1–31. DOI: 10.1007/s00526-012-0538-8.
- [RR17] Gesine Reinert and Nathan Ross. “Approximating stationary distributions of fast mixing Glauber dynamics, with applications to exponential random graphs”. In: *arXiv preprint* (2017). arXiv: 1712.05736v2.
- [Vig01] Eric Vigoda. “A note on the Glauber dynamics for sampling independent sets”. In: *Electron. J. Combin.* 8.1 (2001), Research Paper 8, 8.

ARTHUR SINULIS, FACULTY OF MATHEMATICS, BIELEFELD UNIVERSITY, BIELEFELD, GERMANY
E-mail address: asinulis@math.uni-bielefeld.de