Logarithmic Sobolev inequalities for finite spin systems and applications

HOLGER SAMBALE[*] and ARTHUR SINULIS[†]

Faculty of Mathematics, Bielefeld University, Postfach 100 131, 33501 Bielefeld, Germany.
E-mail: [*]hsambale@math.uni-bielefeld.de; [†]asinulis@math.uni-bielefeld.de

We derive sufficient conditions for a probability measure on a finite product space (a *spin system*) to satisfy a (modified) logarithmic Sobolev inequality. We establish these conditions for various examples, such as the (vertex-weighted) exponential random graph model, the random coloring and the hard-core model with fugacity.

This leads to two separate branches of applications. The first branch is given by mixing time estimates of the Glauber dynamics. The proofs do not rely on coupling arguments, but instead use functional inequalities. As a byproduct, this also yields exponential decay of the relative entropy along the Glauber semigroup. Secondly, we investigate the concentration of measure phenomenon (particularly of higher order) for these spin systems. We show the effect of better concentration properties by centering not around the mean, but around a stochastic term in the exponential random graph model. From there, one can deduce a central limit theorem for the number of triangles from the CLT of the edge count. In the Erdős–Rényi model the first-order approximation leads to a quantification and a proof of a central limit theorem for subgraph counts.

Keywords: central limit theorem; concentration of measure; exponential random graph model; finite product spaces; logarithmic Sobolev inequality; mixing time; spin systems

1. Introduction

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

From the physical viewpoint, a spin system models a collection of particles attaining different states and interacting with each other, so that the complete system consists of a set of *configurations* of the form $\mathcal{X}^{\mathcal{I}}$. Mathematically, a spin system can be described as a probability measure μ on such a product space $\mathcal{Y} := \mathcal{X}^{\mathcal{I}}$, and hard constraints translate into conditions on the support of the probability measure. Here, we consider finite spin systems, that is, the sets $\mathcal{X}$ (the *spins*) and $\mathcal{I}$ (the *sites*) are finite.

Albeit very elementary, these finite spin systems can have a rich dependence structure among the sites. Indeed, many toy models of statistical mechanics are defined as finite spin systems (as the aforementioned Ising model on a finite graph). We are interested in the regimes in which the sites exhibit behavior typical of independent random variables. To this end, we define suitable notions of *weak dependence* which, on the technical side, lead to (modified) logarithmic Sobolev inequalities.

Based on these general results, we study two different branches of applications. Firstly, we want to study the mixing time of the *Glauber dynamics* associated to a sequence of weakly dependent spin systems $(\mu_n)_n$ on $\mathcal{X}^{\mathcal{I}_n}$ which are rapidly mixing, that is, the mixing time is $O(|\mathcal{I}_n| \log|\mathcal{I}_n|)$. Secondly, we study the concentration of measure phenomenon (of higher order) for these kinds of models. From the very definition of both concepts, it is necessary to consider a sequence of finite spin systems with increasing number of sites.

1.1. The models

Our main results are valid for arbitrary weakly dependent spin systems. At the same time, throughout this paper we put a special focus on a number of models for which we establish sufficient conditions for weak dependence and apply our general results. Let us emphasize that any of the models depends on a parameter $n \in \mathbb{N}$, so that we are considering a sequence of spin systems with a growing number of sites. We will usually suppress this dependence on the number of sites. We sometimes specify a spin system by defining a Hamiltonian, that is, a function $H : \mathcal{Y} \to \mathbb{R}$. The spin system associated to H is given by the *Gibbs measure*

$$\mu(\sigma) = \mu_H(\sigma) = Z^{-1} \exp\big(H(\sigma)\big) \quad \text{for } Z = \sum_{\sigma \in \mathcal{Y}} \exp\big(H(\sigma)\big).$$

1.1.1. *The exponential random graph model*

The first spin system we consider is the exponential random graph model. For a thorough historical overview and asymptotic results, we refer to the well-written survey [11]. The model is an extension of the Erdős–Rényi model to account for dependence between the edges.

For two simple graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ let $N_{G_1}(G_2)$ be the number of graph homomorphisms from G_1 to G_2, that is, injective maps $\varphi : V_1 \to V_2$ which preserve edges. Moreover $\mathcal{G}_n$ shall be the set of all simple graphs on n vertices, labeled $\{1, \ldots, n\}$, and $\mathcal{I}_n := \{(i, j) \in \{1, \ldots, n\}^2 : i < j\}$. Let $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_s) \in \mathbb{R}^s$ and $G_1, \ldots, G_s$ be simple, connected graphs $G_i = (V_i, E_i)$. The exponential random graph model with parameters $(\boldsymbol{\beta}, G_1, \ldots, G_s)$, denoted by $\mu_{\boldsymbol{\beta}}$, is defined as the spin system on $\{0, 1\}^{\mathcal{I}_n}$ associated to the Hamiltonian

$$H_{\boldsymbol{\beta}}(x) := n^2 \sum_{i=1}^{s} \beta_i \frac{N_{G_i}(x)}{n^{|V_i|}}.$$

We frequently abbreviate this as $\mathrm{ERGM}(\boldsymbol{\beta}, G_1, \ldots, G_s)$. By convention, G_1 is the complete graph on two vertices K_2. Note that for $s = 1$ we obtain the Erdős–Rényi model with parameter $p = e^{\beta}(1 + e^{\beta})^{-1}$.

For any set of parameters $(\boldsymbol{\beta}, G_1, \ldots, G_s)$, we define the functions $\Phi_{\boldsymbol{\beta}}, \varphi_{\boldsymbol{\beta}} : [0, 1] \to \mathbb{R}$

$$\Phi_{\boldsymbol{\beta}}(y) = \sum_{i=1}^{s} \beta_i |E_i| y^{|E_i|-1} = \beta_1 + \sum_{i=2}^{s} \beta_i |E_i| y^{|E_i|-1},$$

$$\varphi_{\boldsymbol{\beta}}(y) = \frac{\exp(2\Phi_{\boldsymbol{\beta}}(y))}{1 + \exp(2\Phi_{\boldsymbol{\beta}}(y))} = \frac{1}{2}\big(1 + \tanh\big(\Phi_{\boldsymbol{\beta}}(y)\big)\big),$$

and set $|\boldsymbol{\beta}| := (|\beta_1|, \ldots, |\beta_s|)$. The function $\Phi'_{|\boldsymbol{\beta}|}$ (i.e., the derivative of $\Phi_{|\boldsymbol{\beta}|}$) will appear in the condition of weak dependence in this model.

1.1.2. *Vertex-weighted random graph model*

A second model of random graphs – the vertex-weighted exponential random graph model – was recently introduced in [13]. The parameter-space is three-dimensional, that is, $\beta = (\beta_1, \beta_2, p) \in \mathbb{R}^2 \times (0, 1)$, and the model is the spin system on $\mathcal{Y} = \{0, 1\}^n$ defined via the Hamiltonian

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

Note that it resembles the Hamiltonian in the exponential random graph model. On the other hand, it can also be seen as an extension of the Curie–Weiss model on the complete graph with interactions given by a quadratic *and* a cubic form. We define the function

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

Similar to the ERGM, the derivative of φ_β will determine whether the system is weakly dependent.

1.1.3. *Random coloring model*

Given a finite graph $G = (V, E)$ and a set of colors $C = \{1, \ldots, k\}$, the configuration space in the random coloring model is the set of all proper colorings $\Omega_0 \subset C^V$, that is, the set of all $\sigma \in C^V$ such that $\{v, w\} \in E \Rightarrow \varphi_v \neq \varphi_w$, and $\mu = \mu(G, C)$ denotes the uniform distribution on Ω_0.

1.1.4. *Hard-core model with fugacity*

Another model with hard constraints is the hard-core model with *fugacity parameter* λ for some $\lambda > 0$. Given a graph $G = (V, E)$, the hard-core model μ is the spin system on $\mathcal{Y} = \{0, 1\}^V$ satisfying

$$\mu(\sigma) = \begin{cases} Z^{-1} \prod_i \lambda^{\sigma_i} & \sigma \text{ admissible}, \\ 0 & \text{otherwise}. \end{cases}$$

Here, an admissible configuration satisfies $\sigma_v \sigma_w = 0$ for all $\{v, w\} \in E$.

1.2. Outline

The structure of this paper is as follows. We formulate the main results for mixing times in Section 2 (see Theorems 2.1 and 2.2) and prove them in Section 5. The concentration of measure results are given in Section 3, with Section 3.1 containing the concentration inequalities for the

triangle count in the ERGM (Theorem 3.2) as well as the central limit theorem results (Corollary 3.3). Section 3.2 discusses the central limit theorem in the Erdős–Rényi model. These results are applications of Theorems 3.6 and 3.7 given in Section 3.3. We prove all concentration results in Section 6. The proof of the (modified) logarithmic Sobolev inequalities in the models from Section 1.1 can be found in Section 4.

2. Mixing times

For spin systems μ with many sites, one way to sample from these is to use the Markov chain convergence theorem. One canonical Markov chain is the Glauber dynamics, which is a $\mathcal{Y}$-valued ergodic Markov chain $(Y_t)_{t \in \mathbb{N}_0}$ with reversible distribution μ. At each step, it selects a site $i \in \mathcal{I}$ uniformly at random and updates it with the conditional probability given $\overline{x}_i = (x_j)_{j \neq i}$, that is, its transition probability is given by

$$P(x, y) = |\mathcal{I}|^{-1} \sum_{i \in \mathcal{I}} \mu(y_i \mid \overline{x}_i) \prod_{j \neq i} \delta_{x_j = y_j},$$

where δ is the Dirac delta. Under mild conditions, if $(Y_t)_{t \in \mathbb{N}_0}$ is a Markov chain on a finite space $\mathcal{Y}$ with a reversible measure ν, the distribution of Y_t converges to ν. This convergence can be quantified using various metrics between probability measures, and we choose the total variation distance $d_{\mathrm{TV}}(\mu_1, \mu_2) := \sup_{A \subset \mathcal{Y}} |\mu_1(A) - \mu_2(A)| = \frac{1}{2} \sum_{x \in \mathcal{Y}} |\mu_1(x) - \mu_2(x)|$. In the continuous-time case, we define the *mixing time* as

$$t_{\mathrm{mix}} := \inf \left\{ t \in \mathbb{R}_+ : \max_{y \in \mathcal{Y}} d_{\mathrm{TV}}\big(P^t(y, \cdot), \nu\big) \leq e^{-1} \right\}.$$

Here, $P^t(y, \cdot)$ is defined as the distribution of Y_t given $Y_0 = y$.

Our proofs rely on modified logarithmic Sobolev inequalities. Let 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 $\mathrm{Ent}_\mu(f) := \mathbb{E}_\mu f \log f - \mathbb{E}_\mu f \log(\mathbb{E}_\mu f)$ for $f \geq 0$ and the Dirichlet form $\mathcal{E}(f, g) := \mathbb{E}_\mu f(-Lg)$. We say that the triple $(\mathcal{Y}, P, \mu)$ (or in short P, if the space and the measure are clear from the context) satisfies a *modified logarithmic Sobolev inequality* with constant ρ_0 (in short: mLSI(ρ_0)), if for all $f : \mathcal{Y} \to \mathbb{R}_+$ we have

$$\mathrm{Ent}_\mu(f) \leq \frac{\rho_0}{2} \mathcal{E}(f, \log f). \tag{1}$$

The smallest ρ_0 in (1) is called modified logarithmic Sobolev (or entropy) constant; see, for example, [6] and the definition of β in [17]. It is known that the modified logarithmic Sobolev constant can be used to bound the mixing time for the total variation distance of (the distribution of) a Markov semigroup and its trend to equilibrium; see, for example, [6], Theorem 2.4.

We will prove the following theorem, which establishes the mLSI(ρ_0) for the models introduced in Section 1.1, and consequently the exponential decay of the entropy along the Glauber semigroup and the rapid mixing property thereof. Recall that rapid mixing means that the mixing time satisfies $t_{\mathrm{mix}} = O(|\mathcal{I}_n| \log |\mathcal{I}_n|)$.

Theorem 2.1. *The Glauber dynamics satisfies a* mLSI(ρ_0) *with* $\rho_0 \leq C|\mathcal{I}_n|$ *in all following spin systems. The constant C may depend on the parameters of the spin system, but not on n.*

1. *Any exponential random graph model* $\mu_{\boldsymbol{\beta}}$ *such that* $\frac{1}{2}\Phi'_{|\boldsymbol{\beta}|}(1) < 1$.
2. *The vertex-weighted ERGM for* $\beta := (\beta_1, \beta_2, p)$ *satisfying* $\sup_{x \in (0,1)} |\varphi'_\beta(x)| < 1$.
3. *The random coloring model* μ_{G_n} *for any sequence of graphs* $G_n = (V_n, E_n)$ *with uniformly bounded maximum degree* Δ *and* $k \geq 2\Delta + 1$.
4. *The hard-core model with fugacity* λ *on any sequence of graphs* $G_n = (V_n, E_n)$ *with bounded maximum degree* Δ *and* $\lambda < \frac{1}{\Delta - 1}$.

Consequently, the relative entropy decreases exponentially along the Glauber semigroup, that is, for any μ*-density* f *and the Glauber semigroup* $(P_t)_{t \geq 0}$ *we have*

$$\text{Ent}_\mu(P_t f) \leq \exp(-2t/\rho_0) \text{Ent}_\mu(f).$$

Finally, in all the cases, the Glauber dynamics is rapidly mixing.

Let us put the results of Theorem 2.1 into context.

Concerning the exponential random graph model, we are sure that the condition $\frac{1}{2}\Phi'_{|\beta|}(1) < 1$ is not optimal. For $\beta_2, \ldots, \beta_s \geq 0$ is was proven in [4], Theorem 5, that the Glauber dynamics is rapidly mixing whenever there is only one solution a^* to the equation $\varphi_\beta(a) = a$ satisfying $\varphi'_\beta(a^*) < 1$, which is implied by our condition. On the other hand, Theorem 2.1 implies exponential decay of the relative entropy along the Glauber semigroup and also applies to negative parameters. The assumption $\frac{1}{2}\Phi'_{|\beta|}(1) < 1$ is also present in [12], Theorem 6.2. In this parameter region, the authors show convergence of the exponential random graph model to a mixture of Erdős–Renyi graphs in the cut-metric in probability. For more details, we refer to [12].

An easy application shows that for the ERGM with $s = 2$ and a simple, connected graph G_2 on e_2 edges, the condition $|\beta_2| < \binom{e_2}{2}^{-1}$ guarantees rapid mixing.

Rapid mixing for the vertex-weighted exponential random graph model has been established in [13], Theorem 7, under a similar condition as in [4].

The Glauber dynamics for the random coloring model on a sequence of bounded-degree graphs was shown to be rapidly mixing in [21] for $k \geq 2\Delta + 1$ via a path coupling approach. We recover these results using the entropy method. Again, this yields exponential decay of the relative entropy along the Glauber semigroup.

It was shown in [33] 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 associated to the hard-core model with fugacity λ is rapidly mixing. Interestingly, with methods closer to the Bakry–Emery theory and a characterization of Ricci curvature for Markov chains, [16] have shown that the hard-core model with fugacity has a positive Ricci curvature under the assumption $\lambda \leq \frac{1}{\Delta}$, which also implies a mLSI(ρ_0).

Theorem 2.1 is itself an application of the following general theorem. The (rather technical) quantity $\widetilde{\beta}$ and the notion of the interdependence matrix J will be defined in Section 4; see equations (16) and (17).

Theorem 2.2. *Let μ be a spin system on $\mathcal{Y} := \mathcal{X}^{\mathcal{I}}$ for finite sets $\mathcal{X}$ and $\mathcal{I}$. Assume that for some constants $\alpha_1, \alpha_2 > 0$ we have the lower bound on the conditional probabilities*

$$\widetilde{\beta}(\mu) \geq \alpha_1 \tag{2}$$

and an upper bound on some interdependence matrix J

$$\|J\|_{2 \to 2} \leq 1 - \alpha_2. \tag{3}$$

The Glauber dynamics associated to μ satisfies an $\mathrm{mLSI}(2|\mathcal{I}|\alpha_1^{-1}\alpha_2^{-2})$. *As a consequence, given any μ-density f, the density $f_t = P_t f$ ($(P_t)_{t \geq 0}$ is the Glauber semigroup) satisfies*

$$\mathrm{Ent}_\mu(P_t f) \leq \mathrm{Ent}_\mu(f) \exp\left(-\frac{\alpha_1 \alpha_2^2}{2|\mathcal{I}|} t\right). \tag{4}$$

Furthermore, if $(\mu_n)_n$ is a sequence of spin systems on $\mathcal{Y}_n := \mathcal{X}^{\mathcal{I}_n}$ (with $|\mathcal{I}_n| \to \infty$) satisfying (2) *and* (3) *for some n-independent constants α_1, α_2, then the sequence of Glauber dynamics is rapidly mixing.*

Clearly, the second part of Theorem 2.2 is of interest only for $|\mathcal{I}_n| \to \infty$ as $n \to \infty$. In the case of spin systems without hard constraints, we can rephrase the conditions (2), (3). Here, we define $I(\mu) := \min_{i \in \mathcal{I}} \min_{y \in \mathcal{Y}} \mu(y_i \mid \overline{y}_i)$ as the minimal conditional probability.

Corollary 2.3. *Let $(\mu_n)_n$ be a sequence of Gibbs measures on configuration spaces $\mathcal{Y}_n$ induced by Hamiltonians $H_n : \mathcal{Y}_n \to \mathbb{R}$. If $I(\mu_n) \geq \alpha_1$ and $\|J_n\|_{2 \to 2} \leq 1 - \alpha_2$ for some $\alpha_1 \in (0, 1)$, $\alpha_2 \in (0, 1)$ and interdependence matrices J_n, then the (sequence of) Glauber dynamics associated to μ_n is rapidly mixing.*

3. Concentration of measure

Informally, as stated in [31], the *concentration of measure phenomenon* can be described as the phenomenon that a function of n i.i.d. random variables $X_1, \ldots, X_n$ tends to be very close to a deterministic quantity (e.g., its expected value or median), if it is not too sensitive to one of its parameters. The function is usually assumed to be Lipschitz continuous in some sense, depending on a suitably adapted notion of a gradient. In other words, the distribution of any Lipschitz function of independent random variables shows strong (more precisely: sub-Gaussian) concentration properties. For an introduction to the concentration of measure phenomenon and functional inequalities, we refer to the two monographs [23] and [10] and the lecture notes [32].

It is also known that the restriction to Lipschitz functions is not necessary, if one aims at (sub)exponential tails. Already the early works of [7,8] prove L^p norm estimates for polynomials of degree d in Rademacher random variables which grow like $p^{d/2}$ (which can be translated into tail estimates). In the setting of independent sub-Gaussian random variables the Hanson–Wright inequality (see [20,29,36]) gives estimates for quadratic forms. In these cases, exponential tail decay holds even though the Lipschitz condition is not satisfied, which might be regarded as

an extension of the concentration of measure phenomenon beyond the setting of Lipschitz-type functions. This idea has been developed in many different works, such as [1,3,9,22,30,34,35] among others.

The concentration of measure results in this article employ the entropy method. To recall some notions, let $(\mathcal{Y}, \mathcal{A}, \mu)$ be a probability space. An operator $\Gamma : L^\infty(\mu) \to L^\infty(\mu)$ is called a *difference operator* if $|\Gamma(af + b)| = a|\Gamma(f)|$ for all $a > 0$, $b \in \mathbb{R}$. We say that μ satisfies a logarithmic Sobolev inequality with respect to Γ (or in short: a Γ-LSI(σ^2)), if for all bounded and measurable functions $f : \mathcal{Y} \to \mathbb{R}$ we have

$$\text{Ent}_\mu(f^2) \le 2\sigma^2 \int \Gamma(f)^2 \, d\mu. \tag{5}$$

Here, $\text{Ent}_\mu(f)$ is the entropy functional. The smallest $\sigma^2 > 0$ such that (5) holds is known as the logarithmic Sobolev constant. From the two properties of a difference operator, one can infer that a Γ-LSI(σ^2) implies a *Poincaré inequality*

$$\text{Var}_\mu(f) \le \sigma^2 \int \Gamma(f)^2 \, d\mu. \tag{6}$$

To any function f and $i \in \mathcal{I}$, we associate the "local variance" in the ith coordinate

$$\mathfrak{d}_i f(x)^2 = \begin{cases} \dfrac{1}{2} \iint \left(f(\overline{x}_i, y) - f(\overline{x}_i, y') \right)^2 d\mu(y \mid \overline{x}_i) \, d\mu(y' \mid \overline{x}_i) & \text{for } \overline{\mu}_i(\overline{x}_i) > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

Here, $\overline{x}_i = (x_j)_{j \in \mathcal{I} \setminus \{i\}}$ is a generic vector in $\mathcal{X}^{\mathcal{I} \setminus \{i\}}$, $\mu(\cdot \mid \overline{x}_i)$ denotes the conditional probability interpreted as a measure on $\mathcal{X}$, and $\overline{\mu}_i$ is the marginal measure on $\mathcal{X}^{\mathcal{I} \setminus \{i\}}$. More generally, for any $S \subset \mathcal{I}$ we write $\overline{x}_S, \mu(\cdot \mid \overline{x}_S), \overline{\mu}_S$ for the obvious analogues.

One type of difference operator is given by $|\mathfrak{d} f| = (\sum_{i \in \mathcal{I}} (\mathfrak{d}_i f)^2)^{1/2}$, where $|\cdot|$ is the Euclidean norm of a vector. Indeed, if (5) holds for $\Gamma(f) = |\mathfrak{d} f|$, we say that μ satisfies a $\mathfrak{d}$-LSI(σ^2). A second type of difference operator is given by $|\mathfrak{h} f| = (\sum_{i \in \mathcal{I}} (\mathfrak{h}_i f)^2)^{1/2}$ for

$$\mathfrak{h}_i f(x) = \left\| f(\overline{x}_i, y) - f(\overline{x}_i, y') \right\|_{L^\infty(\mu(\overline{x}_i, \cdot) \otimes \mu(\overline{x}_i, \cdot))}. \tag{8}$$

It is easy to see that if μ satisfies a $\mathfrak{d}$-LSI(σ^2), then it also satisfies an $\mathfrak{h}$-LSI$(\sigma^2/2)$.

As a short remark, let us note that the definition of a $\mathfrak{d}$-LSI(σ^2) is consistent with the definition using the Dirichlet form as in Section 2.

Similar to Theorem 2.1, the following theorem provides $\mathfrak{d}$-LSIs for all the models from Section 1.1.

Theorem 3.1. *In the situation of Theorem* 2.1, *the spin systems under the corresponding conditions satisfy a $\mathfrak{d}$-LSI(σ^2), where σ^2 may depend on the parameters of each model but not on n.*

Actually, considering the proof of Theorem 2.1 and Theorem 3.1, one can see that both follow the same strategy: Show that the entropy tensorizes, and apply inequalities which are straightforward to prove. In the case of the logarithmic Sobolev inequality, the constant can be much worse,

as one has an additional dependence on α_1 as given in Theorem 2.2. Aside from this dependence, one can deduce Theorem 2.1 from Theorem 3.1.

3.1. Triangle counts in the exponential random graph model

The question of the distribution of the number of triangles in the Erdős–Rényi model is quite classical and well studied. Therefore, it is an interesting task to find analogous results for the exponential random graph model. Although the edges in this model are dependent, a weak dependence condition should suffice to expect a similar behavior as in the case of independent edges (e.g., see the large deviation results in [12], which in certain cases implies that an exponential random graph model is indistinguishable from an Erdős–Rényi model in the limit). Although large deviation results are purely asymptotic, one can still hope for similar behavior concerning certain statistics for finite n.

Recall that the exponential random graph model is a spin system with sites $\mathcal{I}_n := \{(i, j) \in \{1, \ldots, n\}^2 : i < j\}$. We let $\binom{\mathcal{I}_n}{3}$ be the set of all possibilities of choosing three distinct edges and

$$\mathcal{T}_n := \left\{ \{e, f, g\} \in \binom{\mathcal{I}_n}{3} : e, f, g \text{ form a triangle} \right\}.$$

We define the number of triangles $T_3(x) := \sum_{\{e_1, e_2, e_3\} \in \mathcal{T}_n} x_{e_1} x_{e_2} x_{e_3}$. Our first result are multilevel concentration inequalities for T_3 and a "linear approximation" thereof. Define $\mu_2 := \mathbb{E}_{\mu_\beta} x_e x_f$ (for some edges $e \neq f \in \mathcal{I}_n$, $e \cap f \neq \varnothing$) and $f_1 := \sum_{e \in \mathcal{I}_n} (x_e - \mathbb{E}_{\mu_\beta}(x_e))$. From the definition of the ERGM, it is clear that μ_2 is well-defined.

Theorem 3.2. *Let μ_β be an ERGM satisfying a $\mathfrak{d}$-LSI(σ^2). There exists a constant $C = C(\sigma^2) > 0$ such that for all $t > 0$ we have the multilevel concentration bounds*

$$\mu_\beta\left(|T_3 - \mathbb{E}_{\mu_\beta} T_3| \geq t\right) \leq 2 \exp\left(-\frac{1}{C} \min\left(\left(\frac{t}{n^{3/2}}\right)^{2/3}, \frac{t}{\mu_1 n^{3/2}}, \left(\frac{t}{\mu_2 n^2}\right)^2\right)\right), \qquad (9)$$

$$\mu_\beta\left(|T_3 - \mathbb{E}_{\mu_\beta} T_3 - (n-2)\mu_2 f_1| \geq t\right) \leq 2 \exp\left(-\frac{1}{C} \min\left(\left(\frac{t}{n^{3/2}}\right)^{2/3}, \frac{t}{\mu_1 n^{3/2}}\right)\right). \quad (10)$$

It is interesting to note the effect of subtracting the random variable $(n - 2)\mu_2 f_1$. As the variance of T_3 is of order n^4, a normalization by n^{-2} is necessary to obtain a stable variance, and inequality (9) gives suitable tail estimates. However, the random variable $T_3 - \mathbb{E}_{\mu_\beta} T_3 - (n - 2)\mu_2 f_1$ concentrates on a narrower range, since the variance is of order n^3, and equation (10) yields stretched-exponential tails in this case. In the Erdős–Rényi model, a short calculation shows

$$\mathrm{Var}(T_3) = \binom{n}{3} p^3(1 - p^3) + \frac{1}{2} n(n-1)(n-2)(n-3) p^5(1 - p),$$

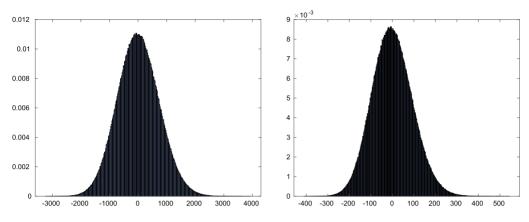$$\mathrm{Var}\left(T_3 - (n-2) p^2 f_1\right) = \binom{n}{3} p^3(1 - p^3).$$

Figure 1. A comparison of the distributions of $T_3 - \mu_\beta(T_3)$ (left) and $T_3 - \mu_\beta(T_3) - (n-2)\mu_2 f_1$ (right) for $n = 100$, $\beta_1 = -0.1$, $\beta_2 = 0.05$ and $G_1 = K_2$ (an edge), $G_2 = K_3$ (a triangle) using the Glauber dynamics and roughly 2 million simulations.

To complement these observations, inspecting (9), we see that the normalization n^{-2} corresponds to the factor n^{-4} in the Gaussian part, whereas the exponential and stretched-exponential part require a normalization of $n^{-3/2}$ only (see Figure 1).

The inequality (10) shows that T_3 fluctuates around the linear term f_1 on a lower order. This leads to the idea of mimicking the method of Hájek projection to deduce a central limit theorem for the triangle count from a CLT for the edge count. As far as we are aware, there are hardly any theoretical results on the distributional limits of the subgraph counts as $n \to \infty$ barring certain special cases. (One such example is the edge two-star model, which can also be interpreted as an Ising model; see [25,26].)

Corollary 3.3. *Let μ_β be an ERGM satisfying a $\mathfrak{d}$-LSI(σ^2). Assuming the central limit theorem $\binom{n}{2}^{-1/2} \sum_{e \in \mathcal{I}_n} (x_e - \mathbb{E}_{\mu_\beta} x_e) \Rightarrow \mathcal{N}(0, v^2)$, we can infer*

$$\frac{T_3 - \mu_\beta(T_3)}{(n-2)\mu_2\sqrt{\binom{n}{2}}} \Rightarrow \mathcal{N}(0, v^2).$$

Actually, the convergence can be quantified in the Wasserstein distance. Let us recall that for two probability measures μ, ν on $\mathbb{R}$ with finite first moment (i.e., $\int |x| \, d\mu(x) < \infty$, $\int |x| \, d\nu(x) < \infty$) the Wasserstein distance is defined as

$$d_W(\mu, \nu) = \sup\left\{ \int f \, d\mu - \int f \, d\nu : f \in \mathrm{Lip}_1 \right\},$$

where Lip_1 denotes the set of all Lipschitz-continuous functions with Lipschitz constant at most 1. For two random variables X, Y, we define $d_W(X, Y)$ as the Wasserstein distance be-

tween its distributions. Define

$$\widetilde{T}_3(x) := (n-2)^{-1}\mu_2^{-1}\binom{n}{2}^{-1/2}\sum_{\{e,f,g\}\in\mathcal{T}_n}\left(x_e x_f x_g - \mathbb{E}_{\mu_\beta}(x_e x_f x_g)\right),$$

$$\widetilde{L}(x) := \binom{n}{2}^{-1/2}\sum_{e\in\mathcal{I}_n}(x_e - \mathbb{E}_{\mu_\beta} x_e).$$

Proposition 3.4. *Let $\mu_\beta = \mu_\beta^{(n)}$ be an ERGM satisfying a $\mathfrak{d}$-LSI(σ^2) and let $Z \sim \mathcal{N}(0, v^2)$ for some $v^2 > 0$. There exists a constant $C = C(\sigma^2)$ such that*

$$d_W(\widetilde{T}_3, Z) \le d_W(\widetilde{L}, Z) + Cn^{-1/2}.$$

Consequently, a rate of convergence in the Wasserstein distance for the number of edges in the ERGM immediately implies a rate of convergence for the number of triangles.

3.2. Central limit theorems for subgraph counts in the Erdős–Rényi model

The second application of the concentration inequalities is a central limit theorem in the classical problem of subgraph counts in the Erdős–Rényi model. Note that the Erdős–Rényi model is a spin system $\mu_{n,p}$ on $\{0,1\}^{\mathcal{I}_n}$ with independent components and $\mu_{n,p}(x_e = 1) = p$.

Define the average degree (or average density) of a graph $d(G) := \max_{H\subset G}\frac{|E(H)|}{|V(H)|}$, and the modified form $d'(G) := \max_{H\subset G, |E(H)|\ge 2}\frac{2|E(H)|-1}{|V(H)|-2}$. Denote by T_G the number of subgraphs G in the Erdős–Rényi random graph, that is, graph homomorphisms from G to the Erdős–Rényi random graph. A possible representation is

$$T_G(X) = \frac{1}{|\mathrm{Aut}(G)|}\sum_{f:V\to[n]\text{ injective}}\prod_{e\in E}X_{f(e)} \tag{11}$$

with the definition $f(e) = \{f(v_1), f(v_2)\}$ for $e = \{v_1, v_2\}$. Moreover, we write $(n)_k := n(n-1)\cdots(n-k+1)$.

Theorem 3.5. *Let $G = (V, E)$ be any simple, connected graph. If $p = p(n)$ satisfies $p \le 1 - \varepsilon$ for some $\varepsilon > 0$ and $np^{d'(G)}\log^{-|E|}(1/p) \to \infty$, then*

$$\frac{T_G - \mu_{n,p}(T_G)}{2(|\mathrm{Aut}(G)|)^{-1}|E|(n-2)_{|V|-2}p^{|E|-1}\sqrt{\binom{n}{2}p(1-p)}} \Rightarrow \mathcal{N}(0, 1).$$

We may replace the condition on $d'(G)$ in Theorem 3.5 by a condition on the maximal degree $d(G)$, though with an additional factor 6, that is, $np^{6d(G)}\log^{-|E|}(1/p) \to \infty$ is a sufficient condition. This is a consequence of the inequality $d(G) \le d'(G) \le 6d(G)$.

Similar calculations (and a proof of a central limit theorem for subgraph counts under nonoptimal conditions) have been done in [27], interpreting subgraph counts as incomplete U-statistics and using the Hoeffding decomposition to prove the CLT. However, [27], Theorem 3.1, does not seem to be quite correct, since for triangles it requires a normalization of the subgraph count by $(n-2)\sqrt{p(1-p)\binom{n}{2}}$, which does not converge to a normal distribution in general. As can be seen from Theorem 3.5, the correct normalization is $(n-2)p^2\sqrt{p(1-p)\binom{n}{2}}$ (see also [15], equation (1.2)). In our approach, we additionally provide a quantification, that is, we show that $T_3 - \mathbb{E}_{\mu_{n,p}} T_3 - p^2(n-2)f_1 = \mathcal{O}_{\mu_{n,p}}(n^{3/2})$ with exponentially decaying tails. We do not believe it is possible to derive similar results using the U-statistics approach.

Using the method of moments, [28] has shown that the convergence holds for any graph G if and only if $np^{d(G)} \to \infty$ and $n^2(1-p) \to \infty$, which is optimal.

3.3. The general results

Theorems 3.2 and 3.5 build upon general concentration properties in the presence of logarithmic Sobolev inequalities. To formulate our results, we introduce higher order differences $\mathfrak{h}_{i_1 \ldots i_d}$ for any $d \in \mathbb{N}$ by setting $\mathfrak{h}_{i_1 \ldots i_d} f = \mathfrak{h}_{i_1}(\mathfrak{h}_{i_2 \ldots i_d} f)$. In particular, we obtain tensors of dth order differences $\mathfrak{h}^{(d)} f$ with coordinates $\mathfrak{h}_{i_1 \ldots i_d} f$. Regarding $\mathfrak{h}^{(d)} f$ as a vector indexed by $\mathcal{I}^d$, we may define $|\mathfrak{h}^{(d)} f|$ as its Euclidean norm. We will write $\|f\|_p$ for the $L^p(\mu)$ norm of f and $\|\mathfrak{h}^{(d)} f\|_p := \||\mathfrak{h}^{(d)} f|\|_p$.

Theorem 3.6. *Let μ be a spin system on $\mathcal{Y} = \mathcal{X}^{\mathcal{I}}$ satisfying a $\mathfrak{d}$-LSI(σ^2). For any $f : \mathcal{Y} \to \mathbb{R}$ and $C := \log(2)\sigma^2(de)^2/2$, we have*

$$\mu\big(|f - \mathbb{E}_\mu f| \geq t\big) \leq 2\exp\left(-\frac{1}{C}\min\left(\frac{t^{2/d}}{\|\mathfrak{h}^{(d)} f\|_\infty^{2/d}}, \min_{k=1,\ldots,d-1}\frac{t^{2/k}}{\|\mathfrak{h}^{(k)} f\|_2^{2/k}}\right)\right). \qquad (12)$$

We apply Theorem 3.6 to functions which resemble multilinear polynomials, constructed in the following way. For any $d \in \mathbb{N}$, we define the (generalized) diagonal of the index set $\mathcal{I}^d$ as $\Delta_d := \{(i_1, \ldots, i_d) \in \mathcal{I}^d : |\{i_1, \ldots, i_d\}| < d\}$. Let $f : \mathcal{X} \to \mathbb{R}$ and A a d-tensor with vanishing diagonal. We can associate to f and any $J \subset \mathcal{I}$ the functions $f_J, \widetilde{f}_J : \mathcal{Y} \to \mathbb{R}$ defined via

$$f_J(y) = \prod_{i \in J} f(y_i) \quad \text{and} \quad \widetilde{f}_i(y) = \prod_{i \in J}\left(f(y_i) - \int f(y_i)\, d\mu\right)$$

and introduce the short-hand notation $\mu_J := \mathbb{E}_\mu f_J$ and $\widetilde{\mu}_J := \mathbb{E}_\mu \widetilde{f}_J$. Given (f, d, A), we construct polynomials as follows: for any finite set $\mathcal{J}$ let

$$\mathcal{P}(\mathcal{J}) = \big\{S \subseteq 2^{\mathcal{J}} : S \text{ is a partition of } \mathcal{J}\big\}.$$

Finally, we set

$$f_{d,A} = \sum_{I \in \mathcal{I}^d} A_I \sum_{P \in \mathcal{P}(I)} g_P := \sum_{I \in \mathcal{I}^d} A_I \sum_{P \in \mathcal{P}(I)} (-1)^{M(P)} \prod_{\substack{J \in P \\ |J|=1}} \widetilde{f}_J \prod_{\substack{J \in P \\ |J|>1}} \widetilde{\mu}_J. \tag{13}$$

Theorem 3.7. *Let μ be a spin system on $\mathcal{Y} = \mathcal{X}^{\mathcal{I}}$ satisfying a $\mathfrak{d}$-LSI(σ^2), $d \in \mathbb{N}$, A a d-tensor with vanishing diagonal and $f : \mathcal{X} \to \mathbb{R}$ with $|f(x) - f(y)| \le c$ for all $x, y \in \mathcal{X}$. Then, $f_{d,A}$ as in (13) is a centered random variable, for all $p \ge 2$ we have*

$$\|f_{d,A}\|_p \le \left(\sigma^2 c^2 \|A\|_2^{2/d} p\right)^{d/2} \tag{14}$$

and consequently

$$\mu\left(|f_{d,A}| \ge t\right) \le 2 \exp\left(-\left(\frac{\log(2)t}{2e\sigma^d c^d \|A\|_2}\right)^{2/d}\right). \tag{15}$$

Note that although (14) seems like an estimate similar to integrability properties of eigenfunctions of the generator L_μ of the Glauber dynamics associated to μ, we do not believe that it can be derived in a similar manner. The reason is that for a nonuniform measure μ the functions $f_{d,A}$ are not eigenfunctions of L_μ, which can easily be seen in the case $d = 1$.

It is possible to refine the tail estimates further by considering different norms of A. As we shall not need it in the present work, we refer to [2] and [18].

4. Weak dependence of the models

We will finally introduce the concept of weakly dependent random variables. Let μ be a spin system on $\mathcal{Y} = \mathcal{X}^{\mathcal{I}}$. Define an *interdependence matrix* $(J_{ij})_{i,j \in \mathcal{I}}$ as any matrix with $J_{ii} = 0$ and such that for any $x, y \in \mathcal{Y}$ with $\overline{x}_j = \overline{y}_j$ we have

$$d_{\mathrm{TV}}\left(\mu(\cdot \mid \overline{x}_i), \mu(\cdot \mid \overline{y}_i)\right) \le J_{ij}. \tag{16}$$

The matrix J (or any norm thereof) may be interpreted as measuring the *strength of the interactions between the spins* in the spin system μ. (In particular, note that $J = 0$ is an interdependence matrix for product measures μ.) Moreover, we need to control the minimal probabilities of the marginal distributions of the spin system μ. To this end, define for any subset $S \subsetneq \mathcal{I}$ and any $i \notin S$

$$\widetilde{\beta}_{i,S}(\mu) := \inf_{\substack{x_S \in \mathcal{X}^S \\ \mu_S(x_S)>0}} \inf_{\substack{y_{S^c} \in \mathcal{X}^{S^c} \\ \mu(y_{S^c}, x_S)>0}} \mu\left((y_{S^c})_i \mid x_S\right). \tag{17}$$

If $S = \varnothing$, this reads $\widetilde{\beta}_{i,\varnothing}(\mu) = \inf_{y \in \mathcal{Y}:\mu(y)>0} \mu(y_i)$. The interpretation of $\widetilde{\beta}_{i,S}(\mu)$ is straightforward: For any admissible partial configuration $x_S \in \mathcal{X}^S$, all possible marginals are supported on points with probability at least $\widetilde{\beta}_{i,S}(\mu)$. Now let

$$\widetilde{\beta}(\mu) := \inf_{S \subsetneq \mathcal{I}} \inf_{i \notin S} \widetilde{\beta}_{i,S}(\mu)$$

be the infimum of all $\widetilde{\beta}_{i,S}(\mu)$. Note that if there are no hard constraints, that is, μ has full support, we have $\widetilde{\beta}(\mu) = I(\mu) = \min_{i \in \mathcal{I}} \min_{y \in \mathcal{Y}} \mu(y_i \mid \overline{y}_i)$.

We say that μ is (α_1, α_2)-*weakly dependent*, if for some interdependence matrix J

$$\widetilde{\beta}(\mu) \geq \alpha_1 \quad \text{and} \quad \|J\|_{2 \to 2} \leq 1 - \alpha_2.$$

Theorem 4.1. *Let μ be a (α_1, α_2)-weakly dependent spin system.*

1. *For any function $f : \mathcal{Y} \to \mathbb{R}_+$ vanishing outside of $\operatorname{supp} \mu$, we have*

$$\operatorname{Ent}_\mu(f) \leq \frac{1}{\alpha_1 \alpha_2^2} \sum_{i \in \mathcal{I}} \int \operatorname{Ent}_{\mu(\cdot \mid \overline{x_i})}\big(f(\overline{x_i}, \cdot)\big) \, d\mu(x). \tag{18}$$

2. *μ satisfies $\mathrm{mLSI}(\rho_0)$ for $\rho_0 := 2|\mathcal{I}|\alpha_1^{-1}\alpha_2^{-2}$.*
3. *μ satisfies $\mathrm{LSI}(\sigma^2)$ for $\sigma^2 := \log(\alpha_1^{-1})(\log(2)\alpha_1\alpha_2^2)^{-1}$.*

Proof of Theorem 4.1. (1): The entropy tensorization property has been established in [24] (see also [19], Theorem 4.2, with a better constant). The condition on the full support of μ can be weakened by adapting the definition of β in [19], Theorem 4.2.

(2): Let us define $\Omega_0 := \operatorname{supp} \mu$, where supp is the support of μ, that is, $\operatorname{supp}(\mu) := \{y \in \mathcal{Y} : \mu(y) > 0\}$. The first part yields for any $f : \mathcal{Y} \to \mathbb{R}$ vanishing outside of Ω_0

$$\operatorname{Ent}_\mu(f) \leq \frac{1}{\alpha_1 \alpha_2^2} \sum_{i \in \mathcal{I}} \int \operatorname{Ent}_{\mu(\cdot \mid \overline{x_i})}\big(f(\overline{x_i}, \cdot)\big) \, d\mu(x). \tag{19}$$

This is equivalent to the fact that on the probability space (Ω_0, μ), any function $f : \Omega_0 \to \mathbb{R}_+$ 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

$$\operatorname{Ent}_\nu\big(e^f\big) \leq \operatorname{Cov}_\nu\big(f, e^f\big) = \int \left(\int \big(f(y) - f(x)\big) \, d\nu(x)\right) e^{f(y)} \, d\nu(y). \tag{20}$$

Apply the inequality (20) to the integral on the right-hand side of equation (19) to get

$$\operatorname{Ent}_\mu\big(e^f\big) \leq \frac{1}{\alpha_1 \alpha_2^2} \sum_{i \in \mathcal{I}} \int \left(\int \big(f(x) - f(\overline{x_i}, y)\big) \, d\mu(y \mid \overline{x_i})\right) e^{f(x)} \, d\mu(x). \tag{21}$$

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

$$\mathcal{E}\big(e^f, f\big) = \mathbb{E}_\mu\big(e^f(-Lf)\big) = \int \sum_{y \in \Omega_0} \big(f(x) - f(y)\big) P(x, y) e^{f(x)} \, d\mu(x)$$

$$= \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} \iint \big(f(x) - f(\overline{x_i}, y)\big) \, d\mu(y \mid \overline{x_i}) e^{f(x)} \, d\mu(x),$$

so that (21) may be rewritten as $\operatorname{Ent}_\mu(e^f) \leq \frac{|\mathcal{I}|}{\alpha_1 \alpha_2^2} \mathcal{E}(e^f, f)$, which yields the claim.

(3): For any $i \in \mathcal{I}$, any $y \in \mathcal{Y}$ with $\mu(y) > 0$ the measure $\mu(\cdot \mid \overline{y}_i)$ is a measure on $\mathcal{X}$ with $\min_{x \in \mathcal{X}} \mu(x \mid \overline{y}_i) \geq \alpha_1$, and so [6], Remark 6.6, yields

$$\mathrm{Ent}_{\mu(\cdot \mid \overline{y}_i)}(g^2) \leq \frac{-2 \log(\alpha_1)}{\log(2)} \mathrm{Var}_{\mu(\cdot \mid \overline{y}_i)}(g),$$

which plugged into equation (18) leads to

$$\mathrm{Ent}_\mu(f^2) \leq 2 \frac{-\log(\alpha_1)}{\log(2)\alpha_1 \alpha_2^2} \sum_{i \in \mathcal{I}} \int \mathrm{Var}_{\mu(\cdot \mid \overline{y}_i)}(f(\overline{y}_i, \cdot)) \, d\mu(y)$$

$$= 2\sigma^2 \int |\eth f(y)|^2 \, d\mu(y). \qquad \square$$

Next, we prove that all the models defined in Section 1.1 satisfy the conditions of Theorem 4.1.

Proposition 4.2. *If β is such that $\frac{1}{2}\Phi'_{|\beta|}(1) < 1$, then μ_β satisfies the conditions of Theorem* 4.1.

Proof of Proposition 4.2. It is convenient to introduce some notation for the exponential random graph model. For any graph $x \in \mathcal{G}_n$ and any edge $e = (i, j) \in \mathcal{I}_n$, let x_{e+} (resp., x_{e-}) be the graph with edge set $E(x_{e+}) = E(x) \cup e$ ($E(x_{e-}) = E(x) \setminus e$ respectively). For any function $f : \mathcal{G}_n \to \mathbb{R}$ we define the discrete derivative in the eth direction as $\eth_e f(x) = f(x_{e+}) - f(x_{e-})$. More generally, given edges $e_1, \ldots, e_k$ we define $\eth_{e_1 \cdots e_k}$ recursively, that is, $\eth_{e_1 \cdots e_k} f = \eth_{e_1}(\eth_{e_2 \cdots e_k} f)$. It is easy to see that the definition does not depend on the order of the edges and $\eth_{ee} f = 0$. The partial derivatives of the Hamiltonian are given by

$$\eth_e H(x) = 2\beta_1 + n^2 \sum_{i=2}^{s} \frac{\beta_i}{n^{|V_i|}} (N_{G_i}(x_{e+}) - N_{G_i}(x_{e-})).$$

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 $\eth_e H(x) = 2\beta_1 + n^2 \sum_{i=2}^{s} \frac{\beta_i}{n^{|V_i|}} N_{G_i}(x, e)$. Especially this gives $|\eth_e H(x)| = O(1)$.

We want to apply Theorem 4.1. 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 $\mathrm{ERGM}(\beta, G_1, \ldots, G_s)$ on $\mathcal{G}_n$. The condition on the conditional distributions is easy to check, since for any $e \in \mathcal{I}_n$ and any $y \in \mathcal{Y}_n$

$$\mu_n(y_e \mid \overline{y}_e) = \frac{1}{2}(1 + \tanh(\eth_e H(y)/2))$$

and $\eth_e H(y) = O(1)$, where the constant depends on $(|\beta|, G_1, \ldots, G_s)$ only. Hence it remains to prove the second condition. 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

$$d_{\mathrm{TV}}\big(\mu_n(\cdot \mid \overline{x}_e), \mu_n(\cdot \mid \overline{y}_e)\big) = \frac{1}{2}\big|\tanh\big(\eth_e H(x_{f+})/2\big) - \tanh\big(\eth_e H(x_{f-})/2\big)\big|$$

$$\leq \frac{1}{4}\big|\eth_{fe} H(x)\big| \leq \frac{n^2}{4}\sum_{i=2}^{s}|\beta_i|\frac{N_{G_i}(x, f, e)}{n^{|V_i|}}$$

$$\leq \frac{n^2}{4}\sum_{i=2}^{s}|\beta_i|\frac{N_{G_i}(K_n, f, e)}{n^{|V_i|}},$$

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

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

Since the right-hand side is independent of $e \in \mathcal{I}_n$, this yields $\|J\|_{1\to 1} \leq \frac{1}{2}\Phi'_{|\beta|}(1) < 1$. Moreover, J is a symmetric matrix, so that we have $\|J\|_{2\to 2} \leq \|J\|_{1\to 1}$. $\square$

Proposition 4.3. *Let μ_β be the vertex-weighted exponential random graph model and assume that $\sup_{\lambda \in (0,1)}|\varphi'_\beta(\lambda)| < 1$. μ_β satisfies the conditions of Theorem 4.1.*

Proof of Proposition 4.3. Since $x_i \in \{0, 1\}$ implies $x_i^k = x_i$ for all $k \in \mathbb{N}$, we can rewrite the Hamiltonian using the order parameter $S := \sum_{i=1}^{n} x_i$ as

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

Hence for $\mathcal{X} := \{0, 1\}$ and $\mathcal{I}_n := \{1, \ldots, n\}$ we are in the situation of Theorem 2.2, and it remains to check conditions (2) and (3). Observe that we have (with the same notation as in the exponential random graph models)

$$\mu(1 \mid \overline{x}_e) = \frac{\exp(\eth_e H_n(\overline{x}_e, 1))}{1 + \exp(\eth_e H_n(\overline{x}_e, 1))} = \frac{1}{2}\big(1 + \tanh\big(\eth_e H_n(x)/2\big)\big),$$

where in this case $|\eth_e H_n(x)| = |\frac{2\beta_1}{n}\sum_{i \neq e} x_i + \frac{3\beta_2}{n^2}\sum_{i \neq j, i, j \neq e} x_i x_j + \log(p/(1-p))|$ is bounded by a constant depending on β, so that a lower bound on the conditional probabilities holds. The inequality (3) is already implicitly proven in the proof of [13], Lemma 6, which we modify. Fix a site $e \in \mathcal{I}_n$ and two configurations x, y differing solely at $f \in \mathcal{I}_n$, that is, $x_f = 1$, $y_f = 0$, and let $S := \sum_{i=1}^{n} y_i$. We have

$$d_{\mathrm{TV}}\big(\mu(\cdot \mid \overline{x}_e), \mu(\cdot \mid \overline{y}_e)\big) = \frac{1}{2}\big|\tanh\big(\eth_e H_n(\overline{x}_e, 1)\big) - \tanh\big(\eth_e H_n(\overline{y}_e, 1)\big)\big|$$

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 $\xi \in (0, 1)$

$$J_{fe} \leq \left| \frac{\exp(h_n((S+1)/n))}{1 + \exp(h_n((S+1)/n))} - \frac{\exp(h_n(S/n))}{1 + \exp(h_n(S/n))} \right| = \frac{1}{n} \left| \left(\frac{\exp \circ h_n}{1 + \exp \circ h_n} \right)'(\xi) \right|. \qquad (22)$$

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 (22) 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

$$\sup_{\lambda \in (0,1)} \left| \frac{\exp \circ h}{1 + \exp \circ h}' \right| < 1 \qquad (23)$$

that (3) holds. $\qquad\square$

Remark. Note that the condition (23) can be written in terms of the functions defined for exponential random graphs. More specifically, we have for any $x \in \mathbb{R}$

$$\frac{\exp(h(x))}{1 + \exp(h(x))} = \varphi_{\widetilde{\beta}_1, \widetilde{\beta}_2, \widetilde{\beta}_3}(x)$$

for the ERGM μ_β given by the three parameters $\widetilde{\beta}_1 = \frac{\log(p/(1-p))}{2}$, $\widetilde{\beta}_2 = \frac{\beta_1}{4}$, $\widetilde{\beta}_3 = \frac{\beta_2}{6}$ and the three graphs G_1 an edge, G_2 a 2-star and G_3 a triangle.

Proposition 4.4. *Let $G_n = (V_n, E_n)$ be a sequence of graphs with uniformly bounded maximum degree Δ and $k \geq 2\Delta + 1$. Then the conditions of Theorem 4.1 hold for the random coloring model μ_{G_n} with α_1, α_2 independent of n.*

Proof of Proposition 4.4. This is again an application of Theorem 4.1. Let us first show that $J_{v,w} := \frac{1}{\Delta + 1} \mathbb{1}_{v \sim w}$ can be used as an interdependence matrix. 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 $\mu_{v_2}(\cdot \mid \overline{c^i}_{v_2})$ are uniform on $C \setminus \{c^i_{v_k} : v_k \sim v_1\}$ for $i = 1, 2$, and hence

$$d_{\mathrm{TV}}\left(\mu_{v_2}(\cdot \mid \overline{c^1}_{v_2}), \mu_{v_2}(\cdot \mid \overline{c^2}_{v_2}) \right)$$

$$= \frac{1}{2}\left(\frac{1}{k - |\{\overline{c^1}_{v_2} : v_2 \sim v_1\}|} + \frac{1}{k - |\{\overline{c^2}_{v_2} : v_2 \sim v_1\}|} \right) \leq \frac{1}{k - \Delta} \leq \frac{1}{\Delta + 1}.$$

On the other hand, if $v_2 \nsim v_1$, then $\mu_{v_2}(\cdot \mid \overline{c^i}_{v_2})$ are equal and thus $J_{v_1,v_2} = 0$. Since J is a symmetric matrix, we obtain

$$\|J\|_{2\to2} \leq \|J\|_{1\to1} \leq \max_{v_j \in V_n} \sum_{v_i \in V_n} J_{v_i,v_j} \leq \frac{\Delta}{\Delta+1} < 1.$$

Moreover, we have to show that $\widetilde{\beta}(\mu_n) \geq \alpha_1$ uniformly in $n \in \mathbb{N}$. Let $S \subsetneq V_n$, $S \neq \varnothing$ be arbitrary, $v_1 \notin S$ and $c^S \in C^S$ be a proper coloring of $G\mid_S = (S, E_n \cap S \times S)$ 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

$$\mu_{\overline{S}}(c^{v_1} \mid c^S) = \frac{\mu(c^{v_1}, c^S)}{\mu(c^S)} = \frac{|\Omega_0(G\mid_S)|}{|\Omega_0(G\mid_{S\cup v_1})|}. \tag{24}$$

It is clear that $|\Omega_0(G\mid_S)| = \frac{1}{|C|}|\Omega_0(\widetilde{G}_S)|$, 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 (24) as follows. Let $N(v_1, S) = \{v_2 \in S : v_2 \sim v_1\} = \{e_1, \ldots, e_l\}$ be the neighbors of v_1 in S and for any $e_1, \ldots, e_k \in N(v_1, S)$ let $G_{e_1,\ldots,e_k}$ be the graph with edge set $(E_n \cap S \times S) \cup \{e_1, \ldots, e_k\}$, so that

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

By [21], equation (2), it follows that each of the ratios is bounded from below by a constant depending on Δ, thus resulting in $\mu_{\overline{S}}(c^{v_1} \mid c^S) \geq |C|^{-1}c(\Delta)$, with a possible choice $c(\Delta) = (\frac{\Delta+1}{\Delta+2})^\Delta$. The case $S = \varnothing$ is easier, as $\mu_i(c^i) = \frac{1}{|C|}$ by the invariance of the random coloring model induced by a relabeling of the colors C. $\qquad\square$

Proposition 4.5. *Let $G = (V, E)$ be any graph with maximum degree Δ. The conditions of Theorem 4.1 hold for the hard core model μ_λ with fugacity λ, and α_1, α_2 depend on Δ only.*

Proof of Proposition 4.5. 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, $\sigma^1, \sigma^2 \in \mathcal{Y}$ 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 $\mu_G(1 \mid \overline{\sigma_{v_1}^1}) = 0$, whereas $\mu_G(1 \mid \overline{\sigma_{v_1}^1}) = \frac{\lambda}{1+\lambda}$. If $v_2 \nsim v_1$, we have $\mu_G(\cdot \mid \overline{\sigma_{v_1}^1}) = \mu_G(\cdot \mid \overline{\sigma_{v_1}^2})$. Hence by the symmetry of J

$$\|J\|_{2\to2} \leq \|J\|_{1\to1} \leq \Delta\frac{\lambda}{1+\lambda} < 1$$

which is a consequence of $\lambda < \frac{1}{\Delta-1}$.

To see that there is a lower bound on the conditional probabilities, let us first consider the case $S = \varnothing$. 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

$$\mu_G(\sigma_v = 1) = \mu(\sigma_v = 1, \sigma_{N(v)} = 0) = Z^{-1} \sum_{\widetilde{\sigma}_A} \mu(\sigma_v = 1, \sigma_{N(v)} = 0, \sigma_A = \widetilde{\sigma}_A)$$

$$= \lambda Z^{-1} \sum_{\widetilde{\sigma}_A} \lambda^{|\widetilde{\sigma}_A|} =: \lambda Z^{-1} Z_A,$$

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

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

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

$$\frac{\lambda}{2^{\Delta+1}} \leq \mu(\sigma_i = 1) \leq \frac{\lambda}{\lambda + 1}. \tag{25}$$

The case $S \neq \varnothing$ follows by a reduction argument. Let $\widetilde{\sigma}_S$ be an admissible configuration of $G \mid_S$ and let $T := \{w \in S : \sigma_w = 0\} \subset S$ be the free sites in S. By explicitly calculating the conditional probability, one can see that for any configuration σ_{S^c} and any $v \notin S$ we have $\mu_G(\sigma_v = 1 \mid \sigma_S = \widetilde{\sigma}_S) = \mu_{G|_{V \setminus T}}(\sigma_v = 1 \mid \sigma_{S \setminus T} = (1, \dots, 1))$. The graph $G \mid_{V \setminus T}$ can be decomposed into three parts: $(T, N(T), R)$, where $N(T) = \bigcup_{v \in T} N(v)$ and

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

which has an upper and lower bound by inequality (25) and so we ultimately get $\widetilde{\beta}(\mu_G) \geq c(\Delta, \lambda)$. □

5. Proofs of the mixing time results

With the results of the last section, the proof of the mixing time results is straightforward.

Proof of Theorem 2.1. Propositions 4.2, 4.3, 4.4 and 4.5 show that all the models satisfy the conditions of Theorem 4.1. Thus, a mLSI(ρ_0) with $\rho_0 \leq C|\mathcal{I}_n|$ holds, where the constant depends on the underlying parameters, but not on n. Hence the assertion follows immediately from Theorem 2.2. □

Proof of Theorem 2.2. The mLSI(ρ_0) property follows immediately from Theorem 4.1. Equation (4) is a consequence of [6], Theorem 2.4, noting that our mLSI constant ρ_0 corresponds to $1/\rho_0$ in [6].

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 $\mu_n^* = \min_{y \in \mathcal{Y}_n} \mu_n(y)$, [6], Corollary 2.8, yields

$$d_{\text{TV}}\left(P^t(y, \cdot), \mu_n\right)^2 \leq 2\log\left(\frac{1}{\mu_n^*}\right)\exp(-2\rho_0^{-1}t) \leq 2\log\left(\frac{1}{\mu_n^*}\right)\exp\left(-\frac{\alpha_1\alpha_2^2}{2|\mathcal{I}_n|}t\right).$$

Hence for $t = \frac{2|\mathcal{I}_n|}{\alpha_1\alpha_2^2} \cdot (\log 2 + 2 + \log\log(1/\mu_n^*))$, we have $\max_{y \in \mathcal{Y}_n} d_{\text{TV}}(P^t(y, \cdot), \mu_n)^2 \leq e^{-2}$, that is, $t_{\text{mix}}(n) \leq \frac{2|\mathcal{I}_n|}{\alpha_1\alpha_2^2} \cdot (\log 2 + 2 + \log\log(1/\mu_n^*))$.

The last step is to show $\log\log 1/\mu_n^* = O(\log|\mathcal{I}_n|)$. This follows using the definition of α_1, since by conditioning and iterating we obtain for any $y \in \mathcal{Y}_n$

$$\frac{1}{\mu_n(y)} = \frac{1}{\mu_n(y_i \mid \overline{y_i})}\mu_n(\overline{y_i})^{-1} \leq \alpha_1^{-1}\mu_n(\overline{y_i})^{-1} \leq \cdots \leq \alpha_1^{-|\mathcal{I}_n|}.$$

Hence, $t_{\text{mix}} = O(|\mathcal{I}_n|\log|\mathcal{I}_n|)$. $\qquad\square$

Proof of Corollary 2.3. If there are no hard constraints, that is, μ has full support, then $\widetilde{\beta}(\mu)$ can be simplified to

$$\widetilde{\beta}(\mu) = I(\mu) := \min_{i \in \mathcal{I}}\min_{y \in \mathcal{Y}}\mu_n(y_i \mid \overline{y_i}).$$

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

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

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

6. Proofs of the concentration of measure results

Proof of Theorem 3.1. The proof is an application of Theorem 4.1, see the proof of Theorem 2.1. $\qquad\square$

We will use Theorem 3.7 to prove several results for finite spin systems. The most important cases of Theorem 3.7 will be $d = 1, 2, 3$. It is easy to check that

$$f_{1,A} := \sum_{i \in \mathcal{I}} A_i \widetilde{f}_i, \qquad f_{2,A} := \sum_{i,j \in \mathcal{I}} A_{ij}(\widetilde{f}_i \widetilde{f}_j - \widetilde{\mu}_{ij}),$$

$$f_{3,A} := \sum_{i,j,k \in \mathcal{I}} A_{ijk}(\widetilde{f_i}\widetilde{f_j}\widetilde{f_k} - \widetilde{\mu}_{ijk} - \widetilde{f_i}\widetilde{\mu}_{jk} - \widetilde{f_j}\widetilde{\mu}_{ik} - \widetilde{f_k}\widetilde{\mu}_{ij}).$$

6.1. Exponential random graph model

We apply the general result on the concentration of the polynomials $f_{d,A}$ from (13) with the spin function $f(x) = x$. Before we prove Theorem 3.2 this way (corresponding to $d = 3$), let us give a simple example which already demonstrates some of the arguments we will use.

Example. Let μ_β be an ERGM satisfying a $\mathfrak{d}$-LSI(σ^2), and let $T_1(x) := \sum_{e \in \mathcal{I}_n} x_e$ be the number of edges. Moreover, for any two disjoint subsets $S_1, S_2 \subset \{1, \ldots, n\}$, write $C(S_1, S_2) := \{e = (i, j) \in \mathcal{I}_n : \{i, j\} \cap S_1 \neq \varnothing, \{i, j\} \cap S_2 \neq \varnothing\}$ and let $T_{S_1, S_2} := \sum_{e \in \mathcal{I}_n} \mathbb{1}_{C(S_1, S_2)}(e)x_e$ be the number of edges between S_1 and S_2. Then

$$\mu_\beta\left(|T_1 - \mathbb{E}_{\mu_\beta(T_1)}| \geq t\right) \leq 2\exp\left(-\frac{\log(2)t^2}{e^2\sigma^2 n(n-1)}\right), \tag{26}$$

$$\mu_\beta\left(\left|T_{S_1, S_2} - \mathbb{E}_{\mu_\beta}(T_{S_1, S_2})\right| \geq t\right) \leq 2\exp\left(-\frac{\log(2)t^2}{2e^2\sigma^2|S_1||S_2|}\right). \tag{27}$$

In particular, setting $\eta := \mathbb{E}_{\mu_\beta}(x_e)$ for $e \in \mathcal{I}_n$ arbitrary, we obtain a strong law of large numbers, that is, $T_1/|\mathcal{I}_n| \to \eta$ a.s.

Proof. Noting that $T_1 - \mathbb{E}_{\mu_\beta}(T_1) = f_{1,A}$ for $A = (1, \ldots, 1)$, (26) readily follows from Theorem 3.7. Equation (27) follows by similar arguments. To prove the strong law of large numbers, first note that by the intrinsic symmetry (i.e., a relabeling of the vertices $\{1, \ldots, n\}$ and a respective relabeling of the edges will result in the same probability law), it is easy to see that $\mu_\beta(x_e)$ does not depend on $e \in \mathcal{I}_n$. Thus, η is well-defined and $\mathbb{E}_{\mu_\beta}(T_1) = |\mathcal{I}_n|\eta$. Now, (26) yields that $T_1/|\mathcal{I}_n|$ converges to η in probability, and the rate of convergence is of order $\exp(-\Omega(n^2))$, which in turn implies convergence almost surely. $\qquad\square$

In a similar vein, we may now prove Theorem 3.2. To this end, we shall express the number of triangles as a linear combination of polynomials of the type $f_{d,A}$.

Proof of Theorem 3.2. For the proof, fix $n \in \mathbb{N}$ and let $X \sim \mu_\beta$. Moreover, we will write $\mu_\Delta := \mathbb{E}X_{e_1 e_2 e_3}$, $\widetilde{\mu}_\Delta := \widetilde{X}_{e_1 e_2 e_3}$ for some triangle $\{e_1, e_2, e_3\}$, $\mu_2 := \mathbb{E}X_{e_1 e_2}$, $\widetilde{\mu}_2 := \mathbb{E}\widetilde{X}_{e_1 e_2}$ for two neighboring edges $e_1 \neq e_2$, $\mu_1 := \mathbb{E}X_e$ and lastly $(e_1, e_2, e_3) \in \mathcal{T}_n$ to indicate that $\{e_1, e_2, e_3\} \in \mathcal{T}_n$ and $e_1 < e_2 < e_3$ (with some fixed partial ordering of the edges).

Now it is not hard to verify that we can decompose $T_3(X) - \mathbb{E}T_3(X)$ as

$$T_3(X) - \mathbb{E}T_3(X) = f_3(X) + \mu_1 f_2(X) + (n-2)\mu_2 f_1(X), \tag{28}$$

using the auxiliary functions

$$f_1(X) := \sum_{e \in \mathcal{I}_n} \widetilde{X}_e, \qquad f_2(X) := \sum_{\substack{e_1 < e_2 \\ e_1 \cap e_2 \neq \varnothing}} (\widetilde{X}_{e_1 e_2} - \widetilde{\mu}_{e_1 e_2}),$$

$$f_3(X) := \sum_{(e_1, e_2, e_3) \in \mathcal{T}_n} (\widetilde{X}_{e_1 e_2 e_3} - \widetilde{\mu}_\Delta - \widetilde{X}_{e_1} \widetilde{\mu}_2 - \widetilde{X}_{e_2} \widetilde{\mu}_2 - \widetilde{X}_{e_3} \widetilde{\mu}_2).$$

Hence, after symmetrization of the sum, the triangle count is the sum of three terms $f_{d,A}$ for different tensors A_3, A_2, A_1 given by $(A_3)_{efg} = 1/6 \cdot \mathbb{1}_{\{e,f,g\} \in \mathcal{T}_n}$, $(A_2)_{ef} = \frac{\mu_1}{2} \mathbb{1}_{e \cap f \neq \varnothing}$ and $(A_1)_e = (n-2)\mu_2$. An easy counting argument shows $\|A_3\|_2 \sim n^{3/2}/6$, $\|A_2\|_2 \sim \mu_1 n^{3/2}/2$ and $\|A_1\|_2 \sim \mu_2 n^2/\sqrt{2}$.

An application of Theorem 3.7 yields

$$\left\| T_3(X) - \mathbb{E}\, T_3(X) \right\|_p \le \left(\sigma^2 \|A_3\|_2^{2/3} p \right)^{3/2} + \left(\sigma^2 \|A_2\|_2 p \right) + \left(\sigma^2 \|A_1\|_2^2 p \right)^{1/2},$$

$$\left\| T_3(X) - \mathbb{E}\, T_3(X) - (n-2)\mu_2 f_1(X) \right\|_p \le \left(\sigma^2 \|A_3\|_2^{2/3} p \right)^{3/2} + \left(\sigma^2 \|A_2\|_2 p \right).$$

The assertion now follows as in the proof of Theorem 3.6. $\qquad\square$

Remark. In the case of the Erdős–Rényi model, the decomposition (28) is simply the Hoeffding decomposition, with $(n-2)\mu_2 f_1(X)$ as the first order and $\mu_1 f_2(X)$ as the second order Hoeffding term, which also coincides with the decomposition of the function $T_3(X)$ in L^2 in terms of the orthonormal basis $(f_S(X))_{S \subset \mathcal{I}_n}$, $f_S(X) = (p(1-p))^{-|S|/2} \prod_{s \in S}(X_s - p)$; see also Section 6.2.

Proof of Corollary 3.3. Using (10), it can be shown that for any $t > 0$,

$$\mu_\beta\left(\left| \frac{T_3 - \mathbb{E}_\mu T_3 - (n-2)\mu_2 f_1}{(n-2)\mu_2 \sqrt{\binom{n}{2}}} \right| \ge t \right) \to 0 \quad \text{for } n \to \infty, \tag{29}$$

and thus

$$\frac{T_3 - \mu_\beta(T_3)}{(n-2)\mu_2 \sqrt{\binom{n}{2}}} = \frac{T_3 - \mu_\beta(T_3) - (n-2)\mu_2 f_1}{(n-2)\mu_2 \sqrt{\binom{n}{2}}} + \frac{1}{\sqrt{\binom{n}{2}}} f_1 \Rightarrow \mathcal{N}(0, \sigma^2)$$

by [5], Theorem 3.1, and the assumption. $\qquad\square$

Remark. Actually equation (29) can be quantified; by (10), the convergence to 0 is of the order $\exp(-\Omega(n^{1/3}))$, which also implies

$$\left((n-2)\mu_2 \binom{n}{2}^{1/2} \right)^{-1} \left(T_3 - \mathbb{E}_\mu T_3 - (n-2)\mu_2 f_1 \right) \to 0 \quad \text{almost surely.}$$

Proof of Proposition 3.4. By the triangle inequality for d_W it suffices to prove that $d_W(\widetilde{T}_3, \widetilde{L}) \leq Cn^{-1/2}$. For any 1-Lipschitz function, we have due to $|f(x) - f(y)| \leq |x - y|$, Theorem 3.2 and a change of variables $s = n^{1/2}t\mu_2$

$$
\left|\mathbb{E}_{\mu_\beta} f(\widetilde{T}_3) - \mathbb{E}_{\mu_\beta} f(\widetilde{L})\right| \leq \int_0^\infty \mu_\beta\left(\left|T_3(x) - (n-2)\mu_2 \sum_{e \in \mathcal{I}_n} \widetilde{x}_e\right| \geq (n-2)\mu_2 \binom{n}{2}^{1/2} t\right) dt
$$

$$
\leq 2 \int_0^\infty \exp\left(-\frac{1}{C} \min\left((\mu_2 n^{1/2} t)^{2/3}, \mu_2 n^{1/2} t\right)\right) dt
$$

$$
\leq 2n^{-1/2}\mu_2^{-1} \int_0^\infty \exp\left(-\frac{1}{C} \min(s^{2/3}, s)\right) ds.
$$

Taking the supremum over all $f \in \mathrm{Lip}_1$ finishes the proof. $\qquad\square$

6.2. Central limit theorems for subgraph counts in Erdös–Rényi graphs

For the Erdős–Rényi model and for $p \in (0, 1)$ we define $\sigma^2(p) := \frac{\log(1-p)-\log(p)}{1-2p}$. Since $\mu_{n,p}$ is a product measure on $\{0, 1\}^{n(n-1)/2}$, by the tensorization property and [14], Theorem A.2, we have

$$
\mathrm{Ent}_{\mu_{n,p}}\left(f^2\right) \leq \sigma^2(p) \int \sum_{i \neq j} \mathrm{Var}_{\mu_{ij}}(f) \, d\mu_n = \sigma^2(p) \int |\partial f|^2 \, d\mu_{n,p}, \tag{30}
$$

that is, $\mu_{n,p}$ satisfies an $\mathrm{LSI}(\sigma^2(p))$. Note that as $p \to 0$ we have $\sigma^2(p) \sim \log(1/p)$, that is, the logarithmic Sobolev constant tends to infinity, however at a logarithmic scale.

Proof of Theorem 3.5. Let $a := (|\mathrm{Aut}(G)|)^{-1}$. We make use of the $L^2(\mu_{n,p})$ (or Hoeffding) decomposition of T_G with respect to the orthonormal basis $(f_S)_{S \subset \mathcal{I}_n}$ given by $f_S = (p(1 - p))^{-|S|/2} \prod_{s \in S}(X_s - p)$, that is,

$$
T_G = \sum_{S \subset \mathcal{I}_n} \langle T_G, f_S \rangle f_S = \sum_{k=0}^{|E|} \sum_{\substack{S \subset \mathcal{I}_n \\ |S|=k}} \langle T_G, f_S \rangle f_S =: \sum_{k=0}^{|E|} T_k = \mu_{n,p}(T_G) + \sum_{k=1}^{|E|} T_k.
$$

For arbitrary $k \in \{1, \ldots, |E|\}$, we obtain from the representation (11),

$$
T_k = a(p(1-p))^{-k} \sum_{\{f_1,\ldots,f_k\}} \sum_{f:V \to [n] \text{ inj}} \left\langle \prod_{e \in E} X_{f(e)}, \widetilde{X}_{f_1 \cdots f_k} \right\rangle \widetilde{X}_{f_1 \cdots f_k}.
$$

For fixed $f_1, \ldots, f_k$, the scalar product is zero unless the injection uses all edges $f_1, \ldots, f_k$, giving

$$T_k = a\big(p(1-p)\big)^{-k} \sum_{\{f_1, \ldots, f_k\}} \widetilde{X}_{f_1 \cdots f_k} \sum_{\substack{f : V \to [n] \text{ inj. uses edges } f_1, \ldots, f_k}} p^{|E|-k}\big(p(1-p)\big)^k$$

$$= a p^{|E|-k} \sum_{\{f_1 \cdots f_k\}} \widetilde{X}_{f_1 \cdots f_k} N_G(f_1, \ldots, f_k),$$

where $N_G(f_1, \ldots, f_k) := |\{f : V \to [n] \text{ inj.} : f \text{ uses edges } f_1, \ldots, f_k\}|$. Especially we have

$$T_1 = \sum_e \langle T_G, f_{\{e\}}\rangle f_{\{e\}} = 2a|E|p^{|E|-1}(n-2)_{|V|-2} \sum_e \widetilde{X}_e.$$

We now claim that

$$\frac{\sum_{k=2}^{|E|} T_k}{2a|E|(n-2)_{|V|-2} p^{|E|-1} \sqrt{\binom{n}{2} p(1-p)}} \to 0 \quad \text{in probability,}$$

from which the result immediately follows, since the normalized first order Hoeffding term converges weakly to a standard normal distribution by the central limit theorem for i.i.d. random variables.

Let us further split the kth Hoeffding term. Denote by $\alpha(f_1, \ldots, f_k)$ the number of vertices that are used in the graph induced by the edge set $\{f_1, \ldots, f_k\}$ and let α_k denote the minimal number of vertices in a subgraph of G with k edges. Clearly, $\alpha(f_1, \ldots, f_k) \in \{\alpha_k, \ldots, 2k \wedge |V|\}$, where $a \wedge b := \min(a, b)$. This results in the decomposition

$$T_k = \sum_{\alpha=\alpha_k}^{|V| \wedge 2k} T_{k,\alpha} \quad \text{and} \quad \widetilde{T}_G := T_G - \mu_{n,p}(T_G) - f_1 = \sum_{k=2}^{|E|} \sum_{\alpha=\alpha_k}^{|V| \wedge 2k} T_{k,\alpha}.$$

Using the $L^q(\mu_{n,p})$ estimates from Proposition 3.7 yields for all $q \geq 2$,

$$\|\widetilde{T}_G\|_q \leq \sum_{k=2}^{|E|} \sum_{\alpha=\alpha_k}^{|V| \wedge 2k} \|T_{k,\alpha}\|_q \leq \sum_{k=2}^{|E|} \sum_{\alpha=\alpha_k}^{|V| \wedge 2k} \big(\sigma^2(p) p^{|E|-k} \|A^{(k,\alpha)}\|_2^{k/2} q\big)^{2/k}.$$

Let $c(n, p) := (|E|(n-2)_{|V|-2} p^{|E|-1} \sqrt{\binom{n}{2} p(1-p)})$. The L^q estimates can be used to show the multilevel concentration inequality (as in the proof of Theorem 3.6)

$$\mu_{n,p}\big(c(n,p)^{-1}|\widetilde{T}_G| \geq t\big) \leq 2\exp\left(-\frac{1}{C\sigma^2(p)} \min_{2 \leq k \leq |E|} \min_{\alpha_k \leq \alpha \leq 2k \wedge |V|} t^{2/k} h_{n,k,\alpha}^{2/k}\right)$$

with $h_{n,k,\alpha} := \frac{c(n,p)}{p^{|E|-k} \|A^{(k,\alpha)}\|_2}$, and it remains to prove that $\min_{k=2,\ldots,|E|} h_{n,k,\alpha} \to \infty$. To this end, we estimate $\|A^{(k,\alpha)}\|_2$ from above as follows. If there are α vertices used by the edges $f_1, \ldots, f_k$,

there are at most $n^{|V|-\alpha}$ ways to have an injection of V into $[n]$ using the edges $f_1, \ldots, f_k$, and there are at most n^α such combinations of $f_1, \ldots, f_k$, and thus

$$\left\| A^{(k,\alpha)} \right\|_2 \leq n^{|V|-\alpha} n^{\alpha/2} = n^{|V|-\alpha/2}.$$

For $k \geq 2$, this gives for some constant $c(G)$ depending on the graph G

$$h_{n,k,\alpha} = \frac{c(n,p)}{p^{|E|-k} \|A^{(k,\alpha)}\|_2} \geq c(G) \frac{n^{|V|-1} p^{|E|-1/2}}{n^{|V|-\alpha/2} p^{|E|-k}} \geq c(G) \left(n^{\alpha-2} p^{2k-1} \right)^{1/2}$$

leading to the condition $\min_{k=2,\ldots,|E|} \min_{\alpha \in \{\alpha_k,\ldots,2k \wedge |V|\}} n p^{(2k-1)/(\alpha-2)} = n p^{d'(G)} \to \infty.$ $\qquad\square$

6.3. Proofs of the general results

Proof of Theorem 3.6. First, note that since μ satisfies a $\mathfrak{d}$-LSI(σ^2), by [19], Proposition 2.4, we obtain for any $p \geq 2$

$$\|f - \mathbb{E}_\mu f\|_p \leq \left(\sigma^2 p\right)^{1/2} \|\mathfrak{h} f\|_p. \tag{31}$$

Next, we iterate (31) using $|\mathfrak{h}|\mathfrak{h}^{(d-1)} f|| \leq |\mathfrak{h}^{(d)} f|$ (cf. [19], Lemma 2.3), leading to

$$\|f - \mathbb{E}_\mu f\|_p \leq \sum_{k=1}^{d-1} \left(\sigma^2 p\right)^{k/2} \left\|\mathfrak{h}^{(k)} f\right\|_2 + \left(\sigma^2 p\right)^{d/2} \left\|\mathfrak{h}^{(d)} f\right\|_p. \tag{32}$$

To prove the multilevel concentration bounds (12), we use methods outlined in [1], Theorem 7 and [3], Theorem 3.3. To sketch the method in a slightly more general situation, assume that there are constants $C_1, \ldots, C_d \geq 0$ such that for any $p \geq 2$

$$\|f - \mathbb{E}_\mu f\|_p \leq \sum_{k=1}^{d} (C_k p)^{k/2}.$$

Let $N := |\{n : C_n > 0\}|$ and $r := \min\{k \in \{1, \ldots, d\} : C_k > 0\}$. By Chebyshev's inequality, we have for any $p \geq 1$

$$\mu\left(|f - \mathbb{E}_\mu f| \geq e\|f - \mathbb{E}_\mu f\|_p\right) \leq \exp(-p). \tag{33}$$

Now consider the function

$$\eta_f(t) := \min\left\{ \frac{t^{2/k}}{C_k} : k = 1, \ldots, d \right\},$$

with $\frac{x}{0}$ being understood as ∞. If we assume $\eta_f(t) \geq 2$, we can estimate $e\|f - \mathbb{E}_\mu f\|_{\eta_f(t)} \leq e \sum_{k=1}^{d} \mathbb{1}_{C_k \neq 0} t = N e t$, so that an application of equation (33) to $p = \eta_f(t)$ yields

$$\mu\left(|f - \mathbb{E}_\mu f| \geq (Ne)t\right) \leq \mu\left(|f - \mathbb{E}_\mu f| \geq e\|f - \mathbb{E}_\mu f\|_{\eta_f(t)}\right) \leq \exp\left(-\eta_f(t)\right).$$

Combining it with the trivial estimate $\mu(\cdot) \leq 1$ (in the case $\eta_f(t) \leq 2$) gives

$$\mu\big(|f - \mathbb{E}_\mu f| \geq (Ne)t\big) \leq e^2 \exp\big(-\eta_f(t)\big).$$

To remove the Ne factor, rescaling the function by Ne and using the estimate $\eta_{(Ne)f}(t) \geq \frac{\eta_f(t)}{(Ne)^{2/r}}$ yields

$$\mu\big(|f - \mathbb{E}_\mu f| \geq t\big) \leq e^2 \exp\Big(-\frac{1}{(Ne)^{2/r}}\eta_f(t)\Big).$$

In a last step, note that $e^2 \exp(-(Ne)^{-2/r}\eta(t)) \leq 2\exp(-\log(2)(Ne)^{-2/r}\eta(t)/2)$ in the nontrivial regime $(Ne)^{-2/r}\eta(t) \geq 2$, so that

$$\mu\big(|f - \mathbb{E}_\mu f| \geq t\big) \leq 2\exp\Big(-\frac{\log(2)}{2(Ne)^{2/r}}\eta_f(t)\Big). \qquad \square$$

To prove Theorem 3.7, let us introduce another notation. For any indices $l_1, \ldots, l_s \in \mathcal{I}$ and s distinct indices $k_1, \ldots, k_s \in \{1, \ldots, d\}$ let $A^{k_1=l_1,\ldots,k_s=l_s}$ be the $(d-s)$-tensor with fixed entries $k_i = l_i$ for all $i = 1, \ldots, s$. For example, if $A = (A_{ijkl})$ is a 4-tensor, $A^{2=j,3=i}$ is the 2-tensor given by $A_{kl}^{2=j,3=i} = A_{kjil}$. Clearly, if A is symmetric, then $A^{k_1=l_1,\ldots,k_s=l_s}$ is symmetric; and if A has a vanishing diagonal, then so has $A^{k_1=l_1,\ldots,k_s=l_s}$.

Proof of Theorem 3.7. To see that $\mathbb{E}_\mu f_{d,A} = 0$ fix $i_1, \ldots, i_d$ and an arbitrary decomposition $P \in \mathcal{P}(\{i_1, \ldots, i_d\})$. If $N(P) = 1$, then g_P has mean zero. On the other hand, if $N(P) \geq 2$, then $P = \{\{i_1\}, \ldots, \{i_{N(P)}\}, I_1, \ldots, I_l\}$ ($l \geq 0$), but the set $\widetilde{P} = \{\{i_1\}, \ldots, i_{N(P)}\}, I_1, \ldots, I_l\}$ is also a valid partition and $g_{\widetilde{P}} = \mathbb{E}_\mu g_P$. As a consequence, $\mathbb{E}_\mu f_{d,A} = 0$.

For any $l \in \mathcal{I}$, write T_l for the formal operator that replaces x_l by $\hat{x}_l$, and use the shorthand notation $x_I := (x_i)_{i \in I}$ for any $I \subset \mathcal{I}$. We shall make use of the inequality

$$\mathfrak{h}_l(f_{d,A}) = \sup_{x_l, \hat{x}_l} \Bigg| \sum_{I=(i_1,\ldots,i_d)} A_I \sum_{P \in \mathcal{P}(I)} (-1)^{M(P)} \big(g_P(x_I) - g_P\big(T_l(x_I)\big)\big) \Bigg|$$

$$= \sup_{x_l, \hat{x}_l} \Bigg| (f(x_l) - f(\hat{x}_l)) \sum_{k=1}^d \sum_{I=(i_1,\ldots,i_{d-1})} A_I^{k=l} \sum_{P \in \mathcal{P}(I)} (-1)^{M(P)} g_P(x_I) \Bigg|$$

$$\leq c \Bigg| \sum_{k=1}^d \sum_{I=(i_1,\ldots,i_{d-1})} A_I^{k=l} \sum_{P \in \mathcal{P}(I)} (-1)^{M(P)} g_P(x_I) \Bigg|$$

$$= c \Big| \sum_{k=1}^d f_{d-1,A^{k=l}} \Big|.$$

Here, the second equality follows from the fact that $T_l(x_{i_1}, \ldots, x_{i_d}) = (x_{i_1}, \ldots, x_{i_d})$ unless $i_j = l$ for some j and the definition of g_P, and the inequality in the third line is a consequence of the assumptions.

We can assume $c = 1$, since the general case follows by rescaling f by c^{-1}. First, by the $\mathfrak{d}$-LSI(σ^2) property we have

$$\|f_{d,A}\|_p^2 \leq \|f_{d,A}\|_2^2 + \sigma^2(p-2)\big\|\mathfrak{h}(f_{d,A})\big\|_p^2.$$

Using the Poincaré inequality with respect to $\mathfrak{h}$ (6) gives

$$\|f_{d,A}\|_2^2 \leq \sigma^2 \sum_{l_1} \mu\big((\mathfrak{h}_{l_1} f_{d,A})^2\big) \leq \sigma^2 \sum_{l_1} \mu\big((\widetilde{\mathfrak{h}_{l_1}} f_{d,A})^2\big) = \sigma^2 \|\widetilde{\mathfrak{h}} f_{d,A}\|_2^2 \leq \sigma^2 \|\widetilde{\mathfrak{h}} f_{d,A}\|_p^2,$$

where $\widetilde{\mathfrak{h}}_l$ replaces $\sup_{x_l, \hat{x}_l} |f(x_l) - f(\hat{x}_l)|$ by 1. Clearly, since $\mathfrak{h}_l f_{d,A} \leq \widetilde{\mathfrak{h}}_l f_{d,A}$ pointwise, the L^p-norms can be estimated as well, resulting in $\|f_{d,A}\|_p^2 \leq \sigma^2(p-1)\|\widetilde{\mathfrak{h}} f_{d,A}\|_p^2$. We have

$$\widetilde{\mathfrak{h}}_{l_1} f_{d,A} = \left| \sum_{k_1=1}^{d} \sum_{I=(i_1,\ldots,i_{d-1})} A_I^{k_1=l_1} \sum_{P \in \mathcal{P}(I)} (-1)^{M(P)} g_P \right|,$$

which itself is the absolute value of a sum of centered random variables, so that the process can be iterated; in each step, the Poincaré inequality (6) can be used and

$$\widetilde{\mathfrak{h}}_{l_1} \cdots \widetilde{\mathfrak{h}}_{l_s} f_{d,A} = \left| \sum_{k_1=1}^{d} \cdots \sum_{k_s=1}^{d-s} \sum_{I=(i_1,\ldots,i_{d-s})} A_I^{k_1=l_1,\ldots,k_s=l_s} \sum_{P \in \mathcal{P}(I)} (-1)^{M(P)} g_P \right|.$$

Thus, using the inequality $\widetilde{h}|\widetilde{h}^{(d)} f| \leq |\widetilde{h}^{(d+1)} f|$ and taking the square root yields

$$\|f_{d,A}\|_p \leq \big(\sigma^2 p\big)^{d/2} \|A\|_2.$$

The multilevel concentration follows as in the proof of Theorem 3.6. $\qquad\square$

Acknowledgments

References

[1] Adamczak, R. (2006). Moment inequalities for U-statistics. *Ann. Probab.* **34** 2288–2314. MR2294982 https://doi.org/10.1214/009117906000000476

[2] Adamczak, R., Kotowski, M., Polaczyk, B. and Strzelecki, M. (2019). A note on concentration for polynomials in the Ising model. *Electron. J. Probab.* **24** Paper No. 42, 22. MR3949267 https://doi.org/10.1214/19-EJP280

[3] Adamczak, R. and Wolff, P. (2015). Concentration inequalities for non-Lipschitz functions with bounded derivatives of higher order. *Probab. Theory Related Fields* **162** 531–586. MR3383337 https://doi.org/10.1007/s00440-014-0579-3

[4] Bhamidi, S., Bresler, G. and Sly, A. (2011). Mixing time of exponential random graphs. *Ann. Appl. Probab.* **21** 2146–2170. MR2895412 https://doi.org/10.1214/10-AAP740

[5] Billingsley, P. (1968). *Convergence of Probability Measures*. New York: Wiley. MR0233396

[6] Bobkov, S.G. and Tetali, P. (2006). Modified logarithmic Sobolev inequalities in discrete settings. *J. Theoret. Probab.* **19** 289–336. MR2283379 https://doi.org/10.1007/s10959-006-0016-3

[7] Bonami, A. (1968). Ensembles $\Lambda(p)$ dans le dual de D^∞. *Ann. Inst. Fourier* (*Grenoble*) **18** 193–204. MR0249940

[8] Bonami, A. (1970). Étude des coefficients de Fourier des fonctions de $L^p(G)$. *Ann. Inst. Fourier* (*Grenoble*) **20** 335–402. MR0283496

[9] Boucheron, S., Bousquet, O., Lugosi, G. and Massart, P. (2005). Moment inequalities for functions of independent random variables. *Ann. Probab.* **33** 514–560. MR2123200 https://doi.org/10.1214/009117904000000856

[10] Boucheron, S., Lugosi, G. and Massart, P. (2013). *Concentration Inequalities*: *A nonasymptotic theory of independence*. Oxford: Oxford Univ. Press. MR3185193 https://doi.org/10.1093/acprof:oso/9780199535255.001.0001

[11] Chatterjee, S. (2016). An introduction to large deviations for random graphs. *Bull. Amer. Math. Soc.* (*N.S.*) **53** 617–642. MR3544262 https://doi.org/10.1090/bull/1539

[12] Chatterjee, S. and Diaconis, P. (2013). Estimating and understanding exponential random graph models. *Ann. Statist.* **41** 2428–2461. MR3127871 https://doi.org/10.1214/13-AOS1155

[13] DeMuse, R., Easlick, T. and Yin, M. (2019). Mixing time of vertex-weighted exponential random graphs. *J. Comput. Appl. Math.* **362** 443–459. MR3987411 https://doi.org/10.1016/j.cam.2018.07.038

[14] Diaconis, P. and Saloff-Coste, L. (1996). Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.* **6** 695–750. MR1410112 https://doi.org/10.1214/aoap/1034968224

[15] Döring, H. and Eichelsbacher, P. (2009). Moderate deviations in a random graph and for the spectrum of Bernoulli random matrices. *Electron. J. Probab.* **14** 2636–2656. MR2570014 https://doi.org/10.1214/EJP.v14-723

[16] Erbar, M., Henderson, C., Menz, G. and Tetali, P. (2017). Ricci curvature bounds for weakly interacting Markov chains. *Electron. J. Probab.* **22** Paper No. 40, 23. MR3646066 https://doi.org/10.1214/17-EJP49

[17] Gao, F. and Quastel, J. (2003). Exponential decay of entropy in the random transposition and Bernoulli–Laplace models. *Ann. Appl. Probab.* **13** 1591–1600. MR2023890 https://doi.org/10.1214/aoap/1069786512

[18] Götze, F., Sambale, H. and Sinulis, A. (2018). Concentration inequalities for bounded functionals via generalized log-Sobolev inequalities. Preprint. Available at arXiv:1812.01092.

[19] Götze, F., Sambale, H. and Sinulis, A. (2019). Higher order concentration for functions of weakly dependent random variables. *Electron. J. Probab.* **24** Paper No. 85, 19. MR4003138

[20] Hanson, D.L. and Wright, F.T. (1971). A bound on tail probabilities for quadratic forms in independent random variables. *Ann. Math. Stat.* **42** 1079–1083. MR0279864 https://doi.org/10.1214/aoms/1177693335

[21] Jerrum, M. (1995). A very simple algorithm for estimating the number of k-colorings of a low-degree graph. *Random Structures Algorithms* **7** 157–165. MR1369061 https://doi.org/10.1002/rsa.3240070205

[22] Latała, R. (2006). Estimates of moments and tails of Gaussian chaoses. *Ann. Probab.* **34** 2315–2331. MR2294983 https://doi.org/10.1214/009117906000000421

[23] Ledoux, M. (2001). *The Concentration of Measure Phenomenon*. *Mathematical Surveys and Monographs* **89**. Providence, RI: Amer. Math. Soc. MR1849347

[24] Marton, K. (2019). Logarithmic Sobolev inequalities in discrete product spaces. *Combin*. *Probab*. *Comput*. **28** 919–935. MR4015662 https://doi.org/10.1017/s0963548319000099

[25] Mukherjee, S. (2013). Phase transition in the two star exponential random graph model. Preprint. Available at arXiv:1310.4164.

[26] Mukherjee, S. (2013). Consistent estimation in the two star exponential random graph model. Preprint. Available at arXiv:1310.4526.

[27] Nowicki, K. and Wierman, J.C. (1988). Subgraph counts in random graphs using incomplete *U*-statistics methods. In *Proceedings of the First Japan Conference on Graph Theory and Applications* (*Hakone*, 1986) **72** 299–310. MR0975550 https://doi.org/10.1016/0012-365X(88)90220-8

[28] Ruciński, A. (1988). When are small subgraphs of a random graph normally distributed? *Probab*. *Theory Related Fields* **78** 1–10. MR0940863 https://doi.org/10.1007/BF00718031

[29] Rudelson, M. and Vershynin, R. (2013). Hanson–Wright inequality and sub-Gaussian concentration. *Electron*. *Commun*. *Probab*. **18** no. 82, 9. MR3125258 https://doi.org/10.1214/ECP.v18-2865

[30] Schudy, W. and Sviridenko, M. (2012). Concentration and moment inequalities for polynomials of independent random variables. In *Proceedings of the Twenty-Third Annual ACM–SIAM Symposium on Discrete Algorithms* 437–446. New York: ACM. MR3205227 https://doi.org/10.1137/1.9781611973099.37

[31] Talagrand, M. (1996). A new look at independence. *Ann*. *Probab*. **24** 1–34. MR1387624 https://doi.org/10.1214/aop/1042644705

[32] van Handel, R. (2016). Probability in high dimension. APC 550 Lecture Notes, Princeton University.

[33] Vigoda, E. (2001). A note on the Glauber dynamics for sampling independent sets. *Electron*. *J*. *Combin*. **8** Research Paper 8, 8. MR1814515

[34] Vu, V.H. (2002). Concentration of non-Lipschitz functions and applications. *Random Structures Algorithms* **20** 262–316. Probabilistic methods in combinatorial optimization. MR1900610 https://doi.org/10.1002/rsa.10032

[35] Wolff, P. (2013). On some Gaussian concentration inequality for non-Lipschitz functions. In *High Dimensional Probability VI*. *Progress in Probability* **66** 103–110. Basel: Birkhäuser/Springer. MR3443496

[36] Wright, F.T. (1973). A bound on tail probabilities for quadratic forms in independent random variables whose distributions are not necessarily symmetric. *Ann*. *Probab*. **1** 1068–1070. MR0353419 https://doi.org/10.1214/aop/1176996815