CONCENTRATION OF MEASURE FOR FINITE SPIN SYSTEMS

HOLGER SAMBALE AND ARTHUR SINULIS

Abstract. In this work we continue the investigation of [GSS18] on *concentration of measure of higher order* for various finite spin systems. We show that under the presence of a logarithmic Sobolev inequality it is possible to estimate the growth of the L^p-norms of any function, which leads to concentration inequalities. Applications to several statistics in the exponential random graph models, the random coloring models, the hard-core model and the Erdös-Renyi model are given. We show the effect of better concentration results by centering not around the mean of the statistic (a zero order approximation), but around a stochastic term (a first order approximation) in the exponential random graph model. In the Erdös-Renyi model we prove a central limit theorem for various subgraph counts.

1. Introduction

The *concentration of measure phenomenon* is by now well understood. Informally, as stated by M. Talagrand in [Tal96], it can be described as the phenomenon that a function of n i.i.d. random variables $X_1, \dots, X_n$ tends to be very close to its expected value, where 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 under product measures shows strong concentration properties.

Here we study a situation in which we usually have weak dependence, and we impose boundedness conditions on the higher order differences of some function f. Moreover, it certain cases it is useful to center f around a certain stochastic expansion to further improve the concentration properties. The spaces under consideration will be finite, though typically large, products of finite spaces, endowed with a measure μ, which we call *spin system*. Examples of spin systems include the Ising model, the exponential random graph model, the random coloring model, the hard-core model and also models with independent entries such as the Erdös-Renyi model.

1.1. **Logarithmic Sobolev inequalities.**

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].

Date: July 21, 2018.

1991 *Mathematics Subject Classification.* Primary 60E, Secondary 60F.

Key words and phrases. concentration of measure, Erdös-Renyi model, exponential random graph model, finite product spaces, logarithmic Sobolev inequalities.

This research was supported by CRC 1283.

We will consider logarithmic Sobolev inequalities on *spin systems*. Let $\mathcal{X}$ and $\mathcal{I}$ be finite sets, and let $\mathcal{Y} = \mathcal{X}^{\mathcal{I}}$ be the set of all configurations and equip $\mathcal{Y}$ with a probability measure μ (called a *spin system*). Define the *entropy functional*

$$\mathrm{Ent}_\mu(f) \coloneqq \mu(f \log f) - \mu(f)\log(\mu(f)) \quad \text{for } f \geq 0 \tag{1.1}$$

and the *Dirichlet form*

$$\begin{aligned}\mathcal{E}_\mu(f,g) &\coloneqq \frac{1}{2}\sum_{i\in\mathcal{I}} \int_{\mathcal{Y}}\int_{\mathcal{X}} (f(\overline{y}_i, y_i) - f(\overline{y}_i, \widetilde{y}))(g(\overline{y}_i, y_i) - g(\overline{y}_i, \widetilde{y}))d\mu(\widetilde{y} \mid \overline{y_i})d\mu(y) \\ &= \sum_{i\in\mathcal{I}} \int \mathrm{Cov}_{\mu(\cdot\mid\overline{y}_i)}(f(\overline{y}_i,\cdot), g(\overline{y}_i,\cdot))d\overline{\mu}_i(\overline{y}_i).\end{aligned} \tag{1.2}$$

Here we use the abbreviation $\mu(f) = \int f d\mu$. Moreover, $\overline{y}_i = (y_j)_{j\in\mathcal{I}\setminus\{i\}}$ is a generic vector, by $\mu(\cdot \mid \overline{y}_i)$ we always mean 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 $\overline{y}_S, \mu(\cdot \mid \overline{y}_S), \overline{\mu}_S$ for the obvious analogues. If $\overline{y}_i$ is such that $\overline{\mu}_i(\overline{y}_i) = 0$ we interpret the integrand as 0.

We say that $(\mathcal{Y}, \mu)$ satisfies a *logarithmic Sobolev inequality* with constant σ^2 (in short $\mathrm{LSI}(\sigma^2)$) if for all $f : \mathcal{Y} \to \mathbb{R}$

$$\mathrm{Ent}_\mu(f^2) \leq 2\sigma^2 \mathcal{E}_\mu(f,f). \tag{1.3}$$

The best constant in (1.3) is known as the logarithmic Sobolev constant. L. Gross [Gro75] was the first to prove the logarithmic Sobolev inequality for the Gaussian measure with another Dirichlet form.

1.2. **Difference operators and concentration of measure.** For any spin system μ satisfying $\mathrm{LSI}(\sigma^2)$ we may obtain concentration of measure results by applying the main result of [GSS18, Theorem 1.5]. To this end, observe that the (diagonal of the) Dirichlet form (1.2) can be rewritten as

$$\mathcal{E}_\mu(f,f) = \int |\mathfrak{d} f|^2 d\mu = \sum_{i\in\mathcal{I}} \int (\mathfrak{d}_i f)^2 d\mu, \tag{1.4}$$

where $|\cdot|$ will always be the Euclidean norm of a vector and

$$\mathfrak{d}_i f(x) = \Big(\frac{1}{2} \int \left(f(x) - f(\overline{x}_i, y)\right)^2 d\mu(y \mid \overline{x}_i) \Big)^{1/2} \tag{1.5}$$

is the “local variance”. This is exactly the difference operator under consideration in [GSS18, Definition 2.2]. A more general definition of a logarithmic Sobolev inequality is with respect to a certain *difference operator*, i.e. an operator $\Gamma : L^\infty(\mu) \to L^\infty(\mu)$ with the property that $|\Gamma(af + b)| = a|\Gamma(f)|$ for all $a > 0, b \in \mathbb{R}$. We say that μ satisfies a logarithmic Sobolev inequality with respect to Γ, if for all $f \in L^\infty(\mu)$ we have

$$\mathrm{Ent}_\mu(f^2) \leq 2\sigma^2 \int \Gamma(f)^2 d\mu. \tag{1.6}$$

Note that the definition above agrees if we define the difference operator $|\mathfrak{d} f| = (\sum_{i\in\mathcal{I}} (\mathfrak{d}_i f)^2)^{1/2}$. 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

$$\mathfrak{h}_i f(x) = \| f(\overline{x}_i, y_i) - f(\overline{x}_i, \widetilde{y}_i)\|_{L^\infty(\mu(\overline{x}_i,\cdot)\otimes\mu(\overline{x}_i,\cdot))}. \tag{1.7}$$

It is easy to see that if μ satisfies $\mathrm{LSI}(\sigma^2)$ with respect to $\mathfrak{d}$, then it also satisfies $\mathrm{LSI}(\sigma^2/2)$ with respect to $\mathfrak{h}$, i.e.

$$\mathrm{Ent}_\mu(f^2) \leq \sigma^2 \int |\mathfrak{h} f|^2 d\mu. \tag{1.8}$$

As usual, this in particular implies a Poincaré inequality

$$\mathrm{Var}_\mu(f) \leq \frac{\sigma^2}{2} \int |\mathfrak{h} f|^2 d\mu.$$

We moreover need to 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 d-th 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 $\|\mathfrak{h}^{(d)} f\|_p := \||\mathfrak{h}^{(d)} f|\|_p$.

A version of [GSS18, Theorem 1.5] adapted to our purposes now reads as follows:

Theorem 1.1. *Let μ be a spin system on $\mathcal{Y} = \mathcal{X}^{\mathcal{I}}$ satisfying $\mathrm{LSI}(\sigma^2)$ with respect to $\mathfrak{d}$, and let $f \in L^\infty(\mu)$. Then for all $p \geq 2$*

$$\|f - \mu(f)\|_p \leq \sum_{k=1}^{d-1} (\sigma^2 p)^{k/2} \|\mathfrak{h}^{(k)} f\|_2 + (\sigma^2 p)^{d/2} \|\mathfrak{h}^{(d)} f\|_p \tag{1.9}$$

In particular, if $\|\mathfrak{h}^{(d)} f\|_\infty \leq C(d,f)$ for some $d \in \mathbb{N}$, then

$$\mu(|f - \mu(f)| \geq t) \leq e^2 \exp\left(- \frac{1}{\sigma^2 (de)^2} \min\left(\frac{t^{2/d}}{C(d,f)^{2/d}}, \min_{k=1,\ldots,d-1} \frac{t^{2/k}}{\|\mathfrak{h}^{(k)} f\|_2^{2/k}} \right) \right). \tag{1.10}$$

Typically, we will apply Theorem 1.1 in the context of subgraph counting. To this end, we will consider functions which resemble multilinear polynomials. For any $d \in \mathbb{N}$ we define the 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 $f : \mathcal{X} \to \mathbb{R}$ be a function which depends on a single spin only, $d \in \mathbb{N}$ and A a d-tensor with vanishing diagonal. We can associate to f the functions $f_i, \widetilde{f}_i : \mathcal{Y} \to \mathbb{R}$ defined via $f_i(y) = f(y_i)$, $\widetilde{f}_i(y) = f(y_i) - \mu_i$, where we use the short-hand notation

$$\mu_{i_1 \ldots i_d} := \int \prod_{j=1}^d f_{i_j} d\mu \tag{1.11}$$

and

$$\widetilde{\mu}_{i_1 \ldots i_d} := \int \prod_{j=1}^d (f_{i_j} - \mu_{i_j}) d\mu \tag{1.12}$$

for any $i_1, \ldots, i_d$. From (f, d, A) we may now construct polynomials as follows: let $\mathcal{J}$ be any set and consider

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

Let $N : \mathcal{P}(\mathcal{J}) \to \mathbb{N}_0$ be the number of singletons in a partition P, i.e. the number of sets $\{i_j\}, i_j \in \mathcal{J}$, and $M : \mathcal{P}(\mathcal{J}) \to \mathbb{N}_0$ the number of subsets with more than one element. To any partition $P \in \mathcal{P}(\mathcal{J})$ we associate a polynomial g_P given by

$$g_P = (-1)^{M(P)} \prod_{\substack{J \in P \\ |J|=1}} \widetilde{f}_J \prod_{\substack{J \in P \\ |J|>1}} \widetilde{\mu}_J.$$

Finally, we set

$$f_{d,A} = \sum_{I=(i_1,\dots,i_d)} A_I \sum_{P \in \mathcal{P}(I)} g_P = \sum_{I=(i_1,\dots,i_d)} A_I \sum_{P \in \mathcal{P}(I)} (-1)^{M(P)} \prod_{\substack{J \in P \\ |J|=1}} \widetilde{f}_J \prod_{\substack{J \in P \\ |J|>1}} \widetilde{\mu}_J. \tag{1.13}$$

The main result of this section is the following concentration inequality for the functions $f_{d,A}$.

Theorem 1.2. *Let μ be a spin system on $\mathcal{Y} = \mathcal{X}^{\mathcal{I}}$ satisfying* $\mathrm{LSI}(\sigma^2)$ *with respect to $\mathfrak{d}$, $d \in \mathbb{N}, A$ 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* (1.13) *is a centered random variable, for all $p \ge 2$ we have*

$$\|f_{d,A}\|_p \le \left(\sigma^2 c^2 \|A\|_2^{2/d} p\right)^{d/2} \tag{1.14}$$

and

$$\mu\left(|f_{d,A}| \ge t\right) \le e^2 \exp\left(-\frac{t^{2/d}}{e^{2/d}\sigma^2 c^2 \|A\|_2^{2/d}}\right). \tag{1.15}$$

Remark. Additionally one can show that also the bound

$$\mu\left(|f_{d,A}| \ge t\right) \le 2 \exp\left(-\frac{t^{2/d}}{4e\sigma^2 c^2 \|A\|_2^{2/d}}\right)$$

holds. This can be done by estimating the exponential moments of $|f_{d,A}|$, see e.g. [BGS17].

If we assume that the spins are independent, then any $P \in \mathcal{P}(I)$ except for $P = \{\{i_1\}, \dots, \{i_d\}\}$ leads to $g_P = 0$. Thus the following Corollary easily follows.

Corollary 1.3. *Assume that $\mu = \otimes_{i=1}^n \mu_i$ is a spin system satisfying* $\mathrm{LSI}(\sigma^2)$ *with respect to $\mathfrak{d}$. For any $d \in \mathbb{N}$, any d-tensor with vanishing diagonal and any function $f : \mathcal{X} \to \mathbb{R}$ with $|f(x) - f(y)| \le c$ we have*

$$\mu\left(\left|\sum_{i_1,\dots,i_d} A_{i_1\dots i_d} \widetilde{f}_{i_1} \cdots \widetilde{f}_{i_d}\right| \ge t\right) \le e^2 \exp\left(-\frac{t^{2/d}}{e^{2/d}\sigma^2 c^2 \|A\|_2^{2/d}}\right)$$

Actually, in the case of independent random variables, the logarithmic Sobolev condition can be significantly weakened or even removed. By the results of S. G. Bobkov, F. Götze and H. Sambale, this can be shown for any set of independent random variables, see [BGS17, Theorem 1.1].

1.3. **Logarithmic Sobolev inequalities for spin systems.** Our results build upon the concept of weakly dependent random variables which enables us to mimic some procedures known from the case of independent random variables. In particular, this includes the tensorization property of the logarithmic Sobolev inequality. The latter was initially proven by K. Marton [Mar15] (see also [GSS18, Theorem 4.1]).

Let μ 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

$$d_{TV}(\mu(\cdot \mid \overline{x}_i), \mu(\cdot \mid \overline{y}_i)) \leq J_{ij}. \tag{1.16}$$

Here, d_{TV} denotes the total variation distance of two measures μ_1, μ_2 on some discrete space X

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

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 if μ is a product measure, then $J \equiv 0$ is an interdependence matrix.

Moreover, we need to control the minimal probabilities of the marginal distributions of the spin system μ. To this end, let for any subset $S \subsetneq \mathcal{I}$

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

If $S = \emptyset$, this reads $\widetilde{\beta}_{i,\emptyset}(\mu) = \inf_{y\in\mathcal{Y}:\mu(y)>0} \mu(y_i)$. 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

$$\widetilde{\beta}(\mu) := \inf_{S \subsetneq \mathcal{I}} \inf_{i \notin S} \widetilde{\beta}_{i,S}(\mu) \tag{1.17}$$

be the infimum of all $\widetilde{\beta}_{i,S}(\mu)$. Note that if there are no hard constraints, i.e. μ has full support, $\widetilde{\beta}(\mu)$ can be simplified to

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

The next theorem establishes a logarithmic Sobolev inequality with a constant depending on $\widetilde{\beta}$ and J for all finite spin systems.

Theorem 1.4. *Let μ be a spin system on $\mathcal{Y} = \mathcal{X}^{\mathcal{I}}$. Assume that for some $\alpha_1, \alpha_2 > 0$*

$$\widetilde{\beta}(\mu) \geq \alpha_1 \qquad \text{and} \qquad \|J\|_{2\to2} \leq 1 - \alpha_2,$$

where J is a suitable interdependence matrix. Then, a logarithmic Sobolev inequality with constant $\sigma^2 := \log(\alpha_1^{-1})(\log(2)\alpha_1\alpha_2^2)^{-1}$ holds, i.e. we have for any $f : \mathcal{Y} \to \mathbb{R}$ vanishing outside of $\mathrm{supp}(\mu)$

$$\mathrm{Ent}_\mu(f^2) \leq 2\sigma^2 \sum_{i\in\mathcal{I}} \frac{1}{2} \int_{\mathcal{Y}} \int_{\mathcal{X}} (f(y) - f(\overline{y}_i, z))^2 d\mu(z \mid \overline{y}_i) d\mu(y). \tag{1.18}$$

Various examples of spin systems satisfying the conditions of Theorem 1.4 will be given in Section 2.

1.4. **Related work.** Tail estimates of the order $t^{2/d}$ for Gaussian chaoses of order d have been proven by R. Latała in [Lat06]. P. Wolff extended this for $d = 2$ to functions of Gaussian random variables with bounded Hessian in [Wol13]. The idea of using higher order derivatives to bound the growth of L^p norms of a function f appears in [Ada06] (for U-statistics) and in [AW15] in the presence of Sobolev-type inequalities. In [GS16] the authors study concentration of *second order* for general independent random variables, and in [BGS17] the applicability is extended to all possible orders, again under independence assumption. In [GSS18], independence has been replaced by the assumption that the measure μ satisfies a logarithmic Sobolev inequality with respect to a probabilistically defined difference operator.

1.5. **Outline.** In Section 2 we show how to use Theorems 1.1 and 1.2 to obtain concentration inequalities for functions of weakly dependent random variables for various models, including the exponential random graph model, the random coloring model, the hard-core model and the Erdös-Renyi model. Section 3 contains the proofs of all results.

2. Applications

We will use Theorem 1.2 to prove several results in finite spin systems. The most important cases of the general result will be $d = 1, 2, 3$. It is easy to check that we have

$$\begin{aligned} f_{1,A} &\coloneqq \sum_{i \in \mathcal{I}} A_i \widetilde{f}_i, \\ f_{2,A} &\coloneqq \sum_{i,j \in \mathcal{I}} A_{ij} \widetilde{f}_i \widetilde{f}_j, \\ f_{3,A} &\coloneqq \sum_{i,j,k \in \mathcal{I}} A_{ijk} \left(\widetilde{f}_i \widetilde{f}_j \widetilde{f}_k - \widetilde{f}_i \widetilde{\mu}_{jk} - \widetilde{f}_j \widetilde{\mu}_{ik} - \widetilde{f}_k \widetilde{\mu}_{ij} \right). \end{aligned}$$

2.1. **Exponential random graph models.** As a first example of a discrete model satisfying a logarithmic Sobolev inequality and thus concentration properties for suitable functionals, we consider the exponential random graph model. To define the model, for any two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ denote by $N_{G_1}(G_2)$ the number of graph homomorphisms $\varphi : G_1 \to G_2$ (i.e. injections $\varphi : V_1 \to V_2$ which preserve edges). Moreover $\mathcal{G}_n$ shall be the set of all simple graphs on n vertices.

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

$$H_{\boldsymbol{\beta}} : \mathcal{G}_n \to \mathbb{R}, \qquad H_{\boldsymbol{\beta}}(x) \coloneqq n^2 \sum_{i=1}^{s} \beta_i \frac{N_{G_i}(x)}{n^{|V_i|}} \tag{2.1}$$

is called Hamiltonian and the probability measure

$$\mu_{\boldsymbol{\beta}}(\{x\}) = Z^{-1} \exp(H_{\boldsymbol{\beta}}(x)) \quad \text{where} \quad Z = \sum_{y \in \mathcal{G}_n} \exp(H_{\boldsymbol{\beta}}(y)) \tag{2.2}$$

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

We will always assume G_1 to be the complete graph on two vertices K_2. For any set of parameters $(\boldsymbol{\beta}, G_1, \ldots, G_s)$ we define the functions $\Phi_{\boldsymbol{\beta}}, \varphi_{\boldsymbol{\beta}} : [0, 1] \to \mathbb{R}$

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

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

For any parameter $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_s)$ we set $|\boldsymbol{\beta}| \coloneqq (|\beta_1|, \ldots, |\beta_s|)$. It is known that the function $\varphi_{\boldsymbol{\beta}}$ determines whether the Glauber dynamics associated to $\mu_{\boldsymbol{\beta}}$ is rapidly mixing or not (see [BBS11], under the assumption of $\beta_i \geq 0$ for all $i = 1, \ldots, s$).

Additionally, under the condition $\frac{1}{2}\Phi'_{|\beta|}(1) < 1$, which also appeared in [CD13, Theorem 6.2], it was shown in [Sin18, Theorem 2.2] that the Glauber dynamics is still rapidly mixing, allowing for possibly negative values of β_i. Especially the conditions of Theorem 1.4 were proven, and thus the following Proposition readily follows.

Proposition 2.2. *Let $\boldsymbol{\beta}$ be such that $\frac{1}{2}\Phi'_{|\beta|}(1) < 1$. Then, $\mu_{\boldsymbol{\beta}}$ satisfies a logarithmic Sobolev inequality with a constant $\sigma^2 = \sigma^2_{\boldsymbol{\beta}}$ depending on $\boldsymbol{\beta}$, i.e.*

$$\operatorname{Ent}_{\mu_{\boldsymbol{\beta}}}(f^2) \le 2\sigma^2 \int |\mathfrak{d} f|^2 d\mu_{\boldsymbol{\beta}}. \tag{2.5}$$

We can use the general result on the concentration of the polynomials $f_{d,A}$ from (1.13) with the spin function $f(x) = x, x \in \mathcal{X} = \{0, 1\}$.

$$\begin{aligned} f_{1,A} &= \sum_{i \in \mathcal{I}_n} A_i \widetilde{x}_i, \\ f_{2,A} &= \sum_{i,j \in \mathcal{I}_n} A_{ij} (\widetilde{x}_{ij} - \widetilde{\mu}_{ij}), \\ f_{3,A} &= \sum_{i,j,k \in \mathcal{I}_n} A_{ijk} \left(\widetilde{x}_{ijk} - \widetilde{\mu}_{ijk} - 3\widetilde{x}_i \widetilde{\mu}_{jk} \right). \end{aligned}$$

In particular, $|f(x) - f(y)| \le 1$, and so the next Corollary immediately follows from Theorem 1.1 and Proposition 2.2.

Corollary 2.3. *Let $\mu_{\boldsymbol{\beta}}$ be an* ERGM *with $\frac{1}{2}\Phi'_{|\beta|}(1) < 1$, $d \in \mathbb{N}$ fixed, A a d-tensor with vanishing diagonal and $f_{d,A}$ as above. We have*

$$\mu_{\boldsymbol{\beta}} \left(|f_{d,A}| \ge t \right) \le e^2 \exp\left(-\frac{t^{2/d}}{e^{2/d} \sigma^2 \|A\|_2^{2/d}} \right) \le e^2 \exp\left(-\frac{t^{2/d}}{e^{2/d} \sigma^2 n(n-1) \|A\|_\infty^{2/d}} \right). \tag{2.6}$$

As can be seen from the proof, the condition $\frac{1}{2}\Phi'_{|\beta|}(1) < 1$ is only needed to make sure that $\mu_{\boldsymbol{\beta}}$ satisfies a logarithmic Sobolev inequality with some constant σ^2 which can be chosen uniformly in $n \in \mathbb{N}$. As a consequence, the results from this section are all valid under the more general assumption of $\mu_{\boldsymbol{\beta}}$ satisfying a logarithmic Sobolev inequality. In particular, proving a logarithmic Sobolev inequality for the full high temperature phase as defined in [BBS11] for $|\boldsymbol{\beta}|$ would extend our results to a region beyond $\frac{1}{2}\Phi'_{|\beta|}(1) < 1$.

2.1.1. *Counting edges.* Our first application is counting the number of edges $T_1 := \sum_{e \in \mathcal{I}_n} x_e$ in models $\mu_{\boldsymbol{\beta}} = \mathrm{ERGM}(\beta, G_1, \ldots, G_s)$ satisfying a logarithmic Sobolev inequality. Note that $T_1 - \mu_{\boldsymbol{\beta}}(T_1) = f_{1,A}$ for $A = (1, \ldots, 1)$. Thus by Corollary 2.3 we obtain

$$\mu_{\boldsymbol{\beta}}(|T_1 - \mu_{\boldsymbol{\beta}}(T_1)| \ge t) \le e^2 \exp\left(-\frac{t^2}{e^2 \sigma^2 |\mathcal{I}_n|} \right) = e^2 \exp\left(-\frac{2t^2}{e^2 \sigma^2 n(n-1)} \right). \tag{2.7}$$

By the intrinsic symmetry (i.e. a relabeling of the vertices $\{1, \ldots, n\}$ and an appropriate relabeling of the edges will result in the same probability law), it is easy to see that $\mu_{\boldsymbol{\beta}}(x_e) =: \eta$ does not depend on $e \in \mathcal{I}_n$. Thus $\mu_{\boldsymbol{\beta}}(T_1) = |\mathcal{I}_n| \eta$.

In particular, (2.7) implies a law of large numbers for the edge count. Indeed, it follows immediately 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.

2.1.2. *Counting edges between two subsets of vertices.* Let $S_1, S_2 \subset \{1, \dots, n\}$ be two disjoint subsets (not necessarily a partition of $\{1, \dots, n\}$). Define

$$C(S_1, S_2) \coloneqq \{e = (i, j) \in \mathcal{I}_n : \{i, j\} \cap S_1 \neq \emptyset, \{i, j\} \cap S_2 \neq \emptyset\}$$

and $A_e \coloneqq \mathbb{1}_{C(S_1,S_2)}(e)$. Clearly we have $|C(S_1, S_2)| = |S_1||S_2|$ and thus $\|A\|_2^2 = \sum_{e\in\mathcal{I}_n} A_e^2 = |S_1||S_2|$. Once again the statistic $T_{S_1,S_2} \coloneqq \sum_{e\in\mathcal{I}_n} A_e x_e$ is concentrated as

$$\mu_{\boldsymbol{\beta}}(|T_{S_1,S_2} - \mu_{\boldsymbol{\beta}}(T_{S_1,S_2})| \geq t) \leq e^2 \exp\left(-\frac{t^2}{e^2\sigma^2|S_1||S_2|}\right). \tag{2.8}$$

2.1.3. *Counting triangles.* Let us define for any $n \in \mathbb{N}$ the set of all triangles

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

Here, $\binom{\mathcal{I}_n}{3}$ denotes the set of all three distinct edges. The number of triangles is given by

$$T_3(x) \coloneqq \sum_{\{e_1,e_2,e_3\}\in\mathcal{T}_n} x_{e_1e_2e_3}. \tag{2.9}$$

To obtain concentration results, we shall express the number of triangles as a linear combination of polynomials of the type $f_{d,A}$. To state our Proposition, we shall also require the function

$$f_1 \coloneqq \sum_{e\in\mathcal{I}_n} \widetilde{x}_e. \tag{2.10}$$

Proposition 2.4. *Let $\mu_{\boldsymbol{\beta}}$ be an exponential random graph model with parameters satisfying $\frac{1}{2}\Phi'_{|\boldsymbol{\beta}|}(1) < 1$. Then, we have*

$$\|T_3 - \mu_{\boldsymbol{\beta}}(T_3)\|_p \leq (\sigma^2 np)^{3/2} + (\sigma^2\mu_1 n^{3/2}p) + (\sigma^2\mu_1^4/2n^4p)^{1/2} \tag{2.11}$$

and

$$\|T_3 - \mu_{\boldsymbol{\beta}}(T_3) - (n-2)\mu_1^2 f_1\|_p \leq (\sigma^2 np)^{3/2} + (\sigma^2\mu_1 n^{3/2}p). \tag{2.12}$$

In particular, this yields the multilevel concentration bounds

$$\mu_{\boldsymbol{\beta}}(|T_3 - \mu_{\boldsymbol{\beta}}(T_3)| \geq t) \leq e^2 \exp\left(-\frac{1}{9e^2\sigma^2}\min\left(\frac{t^{2/3}}{n}, \frac{t}{\mu_1 n^{3/2}}, \frac{2t^2}{\mu_1^4 n^4}\right)\right) \tag{2.13}$$

and

$$\mu_{\boldsymbol{\beta}}\left(|T_3 - \mu_{\boldsymbol{\beta}}(T_3) - (n-2)\mu_1^2 f_1| \geq t\right) \leq e^2 \exp\left(-\frac{1}{2e\sigma^2}\min\left(\frac{t^{2/3}}{n}, \frac{t}{\mu_1 n^{3/2}}\right)\right). \tag{2.14}$$

Remark. It is interesting to note the effect of subtracting the random variable $(n-2)\mu_1^2 f_1$. The variance of T_3 is of order n^4, and thus a normalization of n^{-2} is necessary to obtain a stable variance, and inequality (2.13) gives suitable tail estimates. However, the random variable $T_3 - (n-2)\mu_1^2 f_1$ concentrates on a much narrower range, since the variance is of order n^3, and equation (2.14) yields stretched-exponential tails in this case. In particular, the term $c(n-2)f_1$ serves to remove the n^4 variance term which stems from the covariances.

It is easiest to see this in the case $s = 1$, i. e. $\boldsymbol{\beta} = \beta_1$ and $G_1 = K_2$. Here we obtain a collection of i.i.d. $\{0, 1\}$-random variables $(X_e)_{e\in\mathcal{I}_n}$ with $\mathbb{P}(X_e = 1) = p$ (depending on β_1), the Erdös-Renyi ensemble. Indeed, a simple calculation gives

$$\operatorname{Var}\left(T_3 - (n-2)\mu_1^2 f_1\right) = p^3(1-p^3)\binom{n}{3}.$$

Moreover, inspecting (2.13), 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. This provides another explanation why after subtraction of the "linear term" $\mu_1^2(n-2)f_1$ a normalization by $n^{-3/2}$ is sufficient.

See Figure 1 for a visualization of these observations.

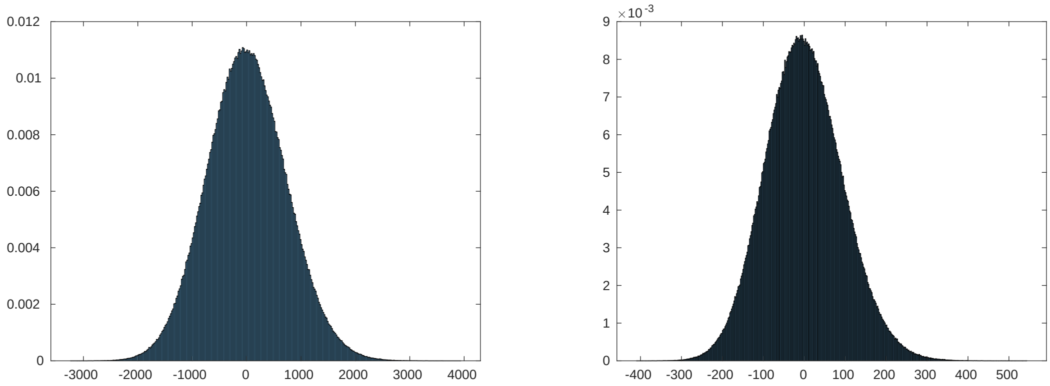

FIGURE 1. A comparison of $T_3 - \mu_\beta(T_3)$ and $T_3 - \mu_\beta(T_3) - (n-2)p^2 f_1$ for $n = 100$, $\beta_1 = -0.1, \beta_2 = 0.05$ and $G_1 = K_2, G_2 = K_3$ using the Glauber dynamics and roughly 2 million simulations.

Proof. We claim that we can decompose T_3 as follows

$$T_3 - \mu(T_3) = f_3 + \mu_1 f_2 + (n-2)\mu_1^2 f_1, \tag{2.15}$$

where we define the auxiliary functions

$$\begin{aligned}
\widetilde{T}_3 &\coloneqq \sum_{\{e_1,e_2,e_3\}\in\mathcal{T}_n} (\widetilde{x}_{e_1e_2e_3} - \widetilde{\mu}_{e_1e_2e_3}) \\
f_3 &\coloneqq \sum_{\{e_1,e_2,e_3\}\in\mathcal{T}_n} (\widetilde{x}_{e_1e_2e_3} - \widetilde{\mu}_{e_1e_2e_3} - 3\widetilde{x}_{e_1}\widetilde{\mu}_{e_2e_3}) \\
f_2 &\coloneqq \sum_{\{e_1,e_2\}:e_1\cap e_2\neq\emptyset} (\widetilde{x}_{e_1e_2} - \widetilde{\mu}_{e_1e_2}).
\end{aligned}$$

To see this, it is easy to see by algebraic manipulations that

$$T_3 - \mu(T_3) = \widetilde{T_3} + r_2 - r_1 \tag{2.16}$$

with

$$\begin{aligned}
r_2 &= 3 \sum_{\{e_1,e_2,e_3\}\in\mathcal{T}_n} (x_{e_1e_2} - \mu_{e_1e_2})\mu_g = \mu_1 \sum_{\{e_1,e_2\}:e_1\cap e_2\neq\emptyset} x_{e_1e_2} - \mu_{e_1e_2} \\
r_1 &= 3 \sum_{\{e_1,e_2,e_3\}\in\mathcal{T}_n} (x_{e_1} - \mu_{e_1})\mu_{e_2e_3} = 3(n-2)\mu_2 f_1
\end{aligned}$$

and it is also immediate to see that

$$f_3 = \widetilde{T_3} - 3(n-2)\widetilde{\mu}_2 f_1. \tag{2.17}$$

Here we have set $\mu_1 := \mu_e$ for some $e \in \mathcal{I}_n$, $\mu_2 := \mu_{ef}$ and $\widetilde{\mu_2} := \mu(\widetilde{x}_{ef})$ for some $e, f \in \mathcal{I}_n$ such that $e \cap f \neq \emptyset$. (Note that these definitions do not depend on the choice of the edges e, f.) Combining equations (2.16) and (2.17) yields

$$T_3 - \mu(T_3) = f_3 + 3(n-2)(\widetilde{\mu}_2 - \mu_2) f_1 + \mu_1 \sum_{\{e_1, e_2\} : e_1 \cap e_2 \neq \emptyset} x_{e_1 e_2} - \mu_{e_1 e_2}. \tag{2.18}$$

Moreover we have

$$f_2 = \sum_{\{e_1, e_2\} : e_1 \cap e_2 \neq \emptyset} (x_{e_1 e_2} - \mu_{e_1 e_2}) - 4\mu_1 (n-2) f_1, \tag{2.19}$$

and thus we can ultimately write

$$\begin{aligned} T_3 - \mu(T_3) &= f_3 + 3(n-2)(\widetilde{\mu}_2 - \mu_2) f_1 + \mu_1 \left(f_2 + 4(n-2)\mu_1 f_1 \right) \\ &= f_3 + \mu_1 f_2 + (n-2) \left(3\widetilde{\mu}_2 - 3\mu_2 + 4\mu_1^2 \right) f_1 \\ &= f_3 + \mu_1 f_2 + (n-2)\mu_1^2 f_1 \end{aligned} \tag{2.20}$$

as claimed.

Hence, after symmetrization of the summands, the triangle count is the sum of three terms $f_{d,A}$ for different tensors, i.e. $f_3 = f_{3,A_3}, f_2 = f_{2,A_2}, f_1 = f_{1,A_1}$ for $(A_3)_{efg} = 1/6 \cdot \mathbb{1}_{\{e,f,g\} \in \mathcal{T}_n}, (A_2)_{ef} = \frac{3}{2}\mu_1 \mathbb{1}_{e \cap f \neq \emptyset}$ and $(A_1)_e = (n-2)\mu_1^2$. The Hilbert-Schmidt norms are

$$\begin{aligned} \|A_3\|_2^2 &= \sum_{(e,f,g) \in \mathcal{T}_n} \frac{1}{36} = \frac{n(n-1)(n-2)}{36} \sim n^3/36 \\ \|A_2\|_2^2 &= \sum_{(e,f) : e \cap f \neq \emptyset} \frac{\mu_1^2}{4} = \mu_1^2 \frac{n(n-1)(n-2)}{4} \sim \mu_1^2 n^3/4 \\ \|A_1\|_2^2 &= \sum_{e \in \mathcal{I}_n} (n-2)^2 \mu_1^4 = (n-2)^2 \mu_1^4 n(n-1)/2 \sim \mu_1^4 n^4/2. \end{aligned}$$

Now it remains to apply Theorem 1.2. $\square$

Remark. In the case of the Erdös-Renyi ensemble, equation (2.15) reduces to the Hoeffding decomposition, with $(n-2)\mu_1^2 f_1$ as the first order and $\mu_1 f_2$ as the second order Hoeffding term, which also coincides with the decomposition of the function T_3 in $L^2(\mu_{n,p})$ in terms of the orthonormal basis $(f_S)_{S \subset \mathcal{I}_n}, f_S = (p(1-p))^{-|S|/2} \prod_{s \in S}(X_s - p)$.

Using the concentration results, we can mimic the method of Hájek projection to show a central limit theorem for the triangle count under the assumption that the number of edges satisfies a central limit theorem. For the special case of the Erdös-Renyi ensemble, we will develop these ideas more in detail in Section 2.4.

Corollary 2.5. *Let μ_β be an exponential random graph model with $\frac{1}{2}\Phi'_{|\beta|}(1) < 1$. If*

$$\frac{1}{\sqrt{\binom{n}{2}}} \sum_{e \in \mathcal{I}_n} (X_e - \mu_e) \Rightarrow \mathcal{N}(0, \sigma^2)$$

holds, then we have

$$\frac{T_3 - \mu_n(T_3)}{(n-2)\mu_1^2 \sqrt{\binom{n}{2}}} \Rightarrow \mathcal{N}(0, \sigma^2).$$

Heuristically, the variance σ^2 should be $p^*(1-p^*)$, where p^* is the unique solution of $p = \varphi_\beta(p), p \in [0,1]$, since in the high temperature phase (with positive β_i) we have $\mathbb{E}\, X_e \to p^*$, see [BBS11, Theorem 7].

Proof. Using the decomposition (2.15) and the concentration of measure results from Proposition 2.4 it can be shown that

$$\mu_\beta \left(\left| \frac{T_3 - \mu(T_3) - (n-2)\mu_1^2 f_1}{(n-2)\mu_1^2 \sqrt{\binom{n}{2}}} \right| \geq t \right) \to 0 \text{ for } n \to \infty, \tag{2.21}$$

and thus

$$\frac{T_3 - \mu_\beta(T_3)}{(n-2)\mu_1^2 \sqrt{\binom{n}{2}}} = \frac{T_3 - \mu_\beta(T_3) - (n-2)\mu_1^2 f_1}{(n-2)\mu_1^2 \sqrt{\binom{n}{2}}} + \frac{1}{\sqrt{\binom{n}{2}}} f_1 \Rightarrow \mathcal{N}(0, \sigma^2)$$

by [Bil68, Theorem 3.1] and the assumption. □

Remark. Actually equation (2.21) can be quantified; by (2.14), the convergence to 0 is of the order $\exp(-\Omega(n^{1/3}))$, and hence $((n-2)\mu_1^2 \binom{n}{2}^{1/2})^{-1} (T_3 - \mu(T_3) - (n-2)\mu_1^2 f_1)$ $\to 0$ almost surely.

2.2. **Random coloring model.** Next, let us consider the random coloring model. Given a graph $G = (V, E)$ and a number of colors $k \geq 2\Delta + 1$, where Δ is the maximum degree of G, the random coloring model μ_G is the measure on $\{1, \ldots, k\}^V$ that assigns equal probability to all proper colorings. The conditions of Theorem 1.4 have been checked in [Sin18, Theorem 2.7] and hence a logarithmic Sobolev inequality (1.3) holds. This especially applies to a sequence of graphs with uniformly bounded degree.

Proposition 2.6. *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 a logarithmic Sobolev inequality holds for μ_{G_n} with a constant $\sigma^2 = \sigma^2(\Delta)$.*

More precisely, given any function $f : \{1, \ldots, k\}^V \to \mathbb{R}$ vanishing outside of $\Omega_0 := \operatorname{supp} \mu = \{(c_v)_{v \in V} \in \{1, \ldots, k\}^V : c_v \neq c_w \ \forall v \sim w\}$ we have

$$\operatorname{Ent}_\mu(f^2) \leq 2\sigma^2 \sum_{c \in \Omega_0} \mu(c) \sum_{v \in V} \int (f(\bar{c}_v, c_v) - f(\bar{c}_v, \tilde{c}_v))^2 d\mu(\tilde{c}_v \mid \bar{c}_v).$$

Similarly, this implies concentration properties for certain functionals. By way of example, we consider the cases of finding vertices, edges and triangles with prescribed colors. We will write the results in terms of a single graph G, but the interesting case will be to consider graphs G_n with $|V_n| \to \infty$ (but with bounded maximal degree Δ_n).

Example (First order statistic). Let $S \subset \{1, \ldots, k\}$. The statistic

$$T_1(c) = \sum_{v \in V} \mathbb{1}_S(c_v),$$

i.e. the number of vertices having a color from S, has the form of $f_{1,A}$ for the function $f(x) = \mathbb{1}_S(x)$ and $A = (1, \ldots, 1)$. Hence by Proposition 1.2 we have

$$\mu_G(|T_1 - \mu_G(T_1)| \geq t) \leq e^2 \exp\left(-\frac{t^2}{e^2 \sigma^2 |V|} \right). \tag{2.22}$$

Note that $\mu_G(T_1) = |V|\frac{|S|}{k}$ by the invariance of the measure μ_G under a relabeling of the colors.

Example (Second order statistic). Again, let $S \subset \{1, \ldots, k\}$ and

$$T_2(c) = \frac{1}{2} \sum_{\substack{v,w \\ v \sim w}} \mathbb{1}_S(c_v)\mathbb{1}_S(c_w) - \mu_G(c_v \in S, c_w \in S)$$

be the number of edges with both endpoints colored with colors from S. If we again choose $f(x) = \mathbb{1}_S(x)$ and the 2-tensor as $A_{vw} = 1/2 \cdot \mathbb{1}_{v \sim w}$, we obtain

$$f_{2,A}(c) = \frac{1}{2} \sum_{\substack{v,w \\ v \sim w}} \widetilde{f}_v \widetilde{f}_w - \widetilde{\mu}_{vw}.$$

It is clear that

$$f_{2,A} = T_2 - \frac{|A|}{k} \sum_{w \in V} \deg(w) \widetilde{f}_w =: T_2 - f_{1,A^{(1)}}. \tag{2.23}$$

From this decomposition, we obtain a two-level bound for the statistic T_2 as

$$\begin{aligned}\|T_2\|_p &\leq \|f_{2,A^{(2)}}\|_p + \|f_{1,A^{(1)}}\|_p \leq \sigma^2 p \|A^{(2)}\|_2 + (\sigma^2 p \|A^{(1)}\|_2^2)^{1/2} \\ &= (\sigma^2 p |E|^{1/2}/2) + \left(\sigma^2 p \frac{|S|^2}{k^2} \sum_{v \in V} \deg(v)^2 \right)^{1/2}.\end{aligned}$$

Using (1.10), we arrive at

$$\mu_G(|T_2| \geq t) \leq e^2 \exp\left(-\frac{1}{4\sigma^2 e^2} \min\left(\frac{2t}{|E|^{1/2}}, \frac{k^2 t^2}{|S|^2 \sum_{v \in V} \deg(v)^2} \right) \right).$$

Example. Lastly, let us consider the case of finding triangles with vertices having prescribed colors. To this end, let $\eta := \frac{|S|}{k} = \mu_G(c_v \in S)$, $\mathcal{T}_n$ be the set of all triangles in G_n and define

$$T_3(c) = \sum_{\{u,v,w\} \in \mathcal{T}_3} \mathbb{1}_S(c_u)\mathbb{1}_S(c_v)\mathbb{1}_S(c_w). \tag{2.24}$$

Again, the centered version can be rewritten

$$T_3 - \mu_G(T_3) = f_{3,A^{(2)}} - 3\eta^2 f_{2,A^{(2)}} - 3\eta f_{1,A^{(1)}}, \tag{2.25}$$

where

$$\begin{aligned} f_{3,A^{(3)}} &= \sum_{\{u,v,w\} \in \mathcal{T}} \widetilde{f}_u \widetilde{f}_v \widetilde{f}_w - \mu_G(\widetilde{f}_u \widetilde{f}_v \widetilde{f}_w) - 3\widetilde{f}_v \mu_G(\widetilde{f}_v \widetilde{f}_w) \\ f_{2,A^{(2)}} &= \sum_{\{u,v\} : u \sim v} \widetilde{f}_u \widetilde{f}_v \Delta(u,v) \\ f_{1,A^{(1)}} &= \sum_u \widetilde{f}_u \Delta(u). \end{aligned}$$

Here $\Delta(u,v) = |S_1(u) \cap S_1(v)|$ is the number of triangles with vertices u and v and $\Delta(u) = |\{v,w\} : \{u,v,w\} \in \mathcal{T}_n|$ is the number of triangles with vertex u. This decomposition yields

$$\mu_G(|T_3 - \mu(T_3)| \geq t) \leq e^2 \exp\left(-\frac{1}{9e^2} \min\left(\frac{t^2}{4\sum_u \Delta(u)^2}, \frac{t}{c|\mathcal{T}_n|^{1/2}}, \frac{t^{2/3}}{|\mathcal{T}_n|^{1/3}} \right) \right). \tag{2.26}$$

Clearly $\sum_u \Delta(u)^2 \le 3\Delta|\mathcal{T}_n|$, and so

$$\mu_G\left(|T_3 - \mu(T_3)| \ge |\mathcal{T}_n|^{1/2} t\right) \le e^2 \exp\left(-\frac{1}{9e^2}\min\left(\frac{t^2}{12\Delta}, \frac{t}{c}, t^{2/3}\right)\right). \tag{2.27}$$

2.3. **Hard-core model.** Another model of a spin system for which a logarithmic Sobolev inequality can be established is the hard-core model with fugacity $\lambda > 0$. Given a graph $G = (V, E)$, it is the probability measure on $\{0,1\}^V$ such that for all admissible configurations σ we have $\mu_\lambda(\sigma) = Z^{-1}\lambda^{\sum_{v\in V}\sigma_v}$. An admissible configuration σ satisfies $\sigma_v\sigma_w = 0$ for all $v \sim w$.

In [Sin18] it was shown that the hard-core model satisfies the conditions of Theorem 1.4 for $\lambda < \frac{1}{\Delta - 1}$ with a constant depending on Δ only.

Proposition 2.7. *Let $G = (V, E)$ be any graph with maximum degree Δ. Then, there exists a constant $\sigma^2 = \sigma^2(\Delta)$ such that* $\mathrm{LSI}(\sigma^2)$ *holds.*

Example. Consider the sequence of cycles C_n (which have uniformly bounded maximum degree $\Delta = 2$), let $\lambda \in (0,1)$ be fixed and denote by μ_n the hard-core model with fugacity λ on C_n. There exists a constant σ^2 so that for all $n \ge 2$ the measure μ_n satisfies $\mathrm{LSI}(\sigma^2)$. If we write the vertices of C_n as $x \in \mathbb{Z}/n\mathbb{Z}$, we can consider the function

$$T_2(\sigma) = n^{-1/2}\sum_{k\in\mathbb{Z}/n\mathbb{Z}} \sigma_k\sigma_{k+2},$$

which counts the number of particles with a distance of 2. A short calculation yields $\mathfrak{h}_e T_2 \le n^{-1/2}$, and from Theorem 1.1 (with $d = 1$) we obtain

$$\mu_n\left(|T_2 - \mu_n(T_2)| \ge t\right) \le e^2 \exp\left(-\frac{t^2}{e^2\sigma^2}\right).$$

It is possible to generalize this example to sequences of graphs $G = G_n$ with uniformly bounded maximum degree Δ. As above, we count the number of particles with a distance of 2. Here we may proceed as above though with additional factors depending on Δ. Indeed, we always have $\mathfrak{h}_e T_2 \le \Delta(\Delta-1)n^{-1/2}$, where equality can only be reached for graphs having a tree as a subgraph.

2.4. **A central limit theorem for Erdös-Renyi graphs.** For any $n \in \mathbb{N}$ and $p = p(n)$ denote by $\mu_{n,p}$ the Erdös-Renyi model on n vertices, i.e. $\mu_{n,p}$ is the product measure of Bernoulli(p) distributions on the $n(n-1)/2$ possible edges. Write

$$\sigma^2(p) = \frac{1-2p}{\log(1-p) - \log(p)}.$$

Since $\mu_{n,p}$ is a product measure on $\{0,1\}^{n(n-1)/2}$, by the tensorization property and [DS96, Theorem A.2] we have

$$\mathrm{Ent}_{\mu_{n,p}}(f^2) \le \sigma^2(p)\int\sum_{i\neq j}\mathrm{Var}_{\mu_{ij}}(f)d\mu_n = \sigma^2(p)\int|\mathfrak{d}f|^2 d\mu_{n,p}. \tag{2.28}$$

Thus $\mu_{n,p}$ satisfies $\mathrm{LSI}(\sigma^2(p))$. Note that as $p \to 0$ we have $\sigma^2(p) \sim (\log(1/p))^{-1}$, i.e. the logarithmic Sobolev constant tends to infinity, however at a logarithmic scale (in p). It is also possible to relate the logarithmic Sobolev constant $\sigma^2(p)$ with a logarithmic Sobolev inequality with respect to the difference operator $\mathfrak{h}$, since

$$\mathfrak{d}_i f(y)^2 = \frac{1}{2}\iint (f(\overline{y}_i, y') - f(\overline{y}_i, y''))^2 d\mu^2(y', y'') = p(1-p)(f(\overline{y}_i, 1) - f(\overline{y}_i, 0))^2$$

$$= p(1-p)\mathfrak{h}_i f(y)^2.$$

Now it is easily seen that equation (2.28) implies

$$\mathrm{Ent}_{\mu_{n,p}}(f^2) \le \sigma^2(p)p(1-p)\int |\mathfrak{h} f|^2 d\mu_{n,p}. \tag{2.29}$$

Using this logarithmic Sobolev inequality, it is possible to prove a central limit theorem for subgraph counts under decay conditions on the sequence p_n. We will first consider the case of finding a triangle to convey the idea, and prove the general theorem thereafter. Recall the definition (2.9) of the triangle count T_3.

Proposition 2.8. *Let $(\mu_{n,p})_n$ be a sequence of Erdös-Renyi models with $p = p(n)$ satisfying*

$$\frac{np^2}{\log^3(1/p)} \to \infty \qquad \textit{and} \qquad p \le 1-\varepsilon \textit{ for some } \varepsilon > 0.$$

Then

$$\frac{T_3 - \mu_{n,p}(T_3)}{(n-2)p^2\sqrt{p(1-p)n(n-1)/2}} \Rightarrow \mathcal{N}(0,1).$$

Remark. The condition is fulfilled if $p(n) = n^{-\alpha}$ and $\alpha \in (0, 1/2)$ (or clearly if $p > \varepsilon$ uniformly in n for some $\varepsilon > 0$). For the triangle count we are off by a p factor (ignoring the logarithmic dependence on p), since in [Ruc88] it has been shown that the convergence holds for any graph G if and only if $np^m \to \infty$, where $m = \max\{e(H)/|H| : H \subset G\}$, which in the case of the k-cycles is 1. However, the case of triangles is the worst case in the set of all cycles as will be apparent from the next proposition.

Proof. Firstly, define $h_n := (n-2)p^2\sqrt{p(1-p)n(n-1)/2}$ and $\widetilde{T}_3 := T_3 - \mu_{n,p}(T_3) - (n-2)p^2 f_1$. Now, applying Proposition 2.4 (or rather Theorem 1.2 and the same calculation as done for Proposition 2.4), we obtain

$$\begin{aligned}\mu_{n,p}\left(\left|\widetilde{T}_3\right| \ge h_n t\right) &\le e^2 \exp\left(-\frac{1}{2e\sigma^2(p)}\min\left(\left(\frac{th_n}{(p(1-p))^{3/2}n^{3/2}}\right)^{2/3}, \frac{th_n}{pn^{3/2}p(1-p)}\right)\right) \\ &\le e^2\exp\left(-\frac{1}{2e\sigma^2(p)(1-p)}\min\left(c_1t^{2/3}(np^2)^{1/3}, c_2t(np)^{1/2}\right)\right).\end{aligned}$$

By assumption, $np^2/(\sigma^2(p))^3 \to \infty$, and we obtain $h_n^{-1}\widetilde{T}_3 \to 0$ in probability (or almost surely, if the divergence of is fast enough).

Secondly, note that

$$\frac{(n-2)p^2}{h_n}\sum_{e\in\mathcal{I}_n}(x_e - p) = \frac{1}{\sqrt{|\mathcal{I}_n|}}\sum_{e\in\mathcal{I}_n}\frac{x_e - p}{\sqrt{p(1-p))}} \Rightarrow \mathcal{N}(0,1)$$

by the central limit theorem for i.i.d. random variables. The Proposition now follows from [Bil68, Theorem 3.1]. □

To prove the general theorem, denote by T_G the number of subgraph counts, i.e. graph homomorphisms from G to the Erdös-Renyi random graph. A possible representation is

$$T_G(X) = \sum_{f:V\to[n] \text{ injective}}\prod_{e\in E} X_{f(e)} \tag{2.30}$$

with the definition $f(e) = \{f(e_1), f(e_2)\}$ for $e = \{e_1, e_2\}$. For notational convenience, we will (unlike as in the triangle case) not divide by the number of automorphisms of G, resulting in $|\mathrm{Aut}(G)|$ in the denominator in the next theorem.

Lastly, define the maximal degree of a graph as $d(G) := \max_{H \subset G} \frac{|E(H)|}{|V(H)|}$, and the modified form $\widetilde{d}(G) := \max_{H \subset G, |E(H)| \geq 2} \frac{|E(H)|-1}{|V(H)|-2}$.

Theorem 2.9. *Let $G = (V, E)$ be any simple, connected graph and let $p = p(n)$ satisfy $np^{\widetilde{d}(G)} / \log^{|E|}(1/p) \to \infty$ and $p \leq 1 - \varepsilon$ for some $\varepsilon > 0$. Then, we have*

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

In particular, if $G = C_m$ is a cycle, we have $\widetilde{d}(C_m) = (m-1)/(m-2)$, which gives back Proposition 2.8 if $m = 3$, while $\widetilde{d}(C_m)$ approaches the optimal value 1 for large m.

Proof. We will consider the $L^2(\mu_{n,p})$ (or Hoeffding) decomposition of T_G with respect to the orthonormal basis $(f_S)_{S \subset \mathcal{I}_n}$, $f_S = (p(1-p))^{-|S|/2} \prod_{s \in S}(X_s - p)$, i.e.

$$T_G = \sum_{S \subset \mathcal{I}_n} \langle T_G, f_S \rangle f_S = \sum_{k=0}^{|E|} \sum_{S \subset \mathcal{I}_n : |S| = k} \langle T_G, f_S \rangle f_S =: \sum_{k=0}^{|E|} T_k.$$

It is clear that $T_0 = \mu_{n,p}(T_G)$ and an easy calculation yields

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

For arbitrary $k \in \{1, \ldots, |E|\}$, this can be seen by considering the representation (2.30): for any distinct edges $f_1, \ldots, f_k$ we obtain

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

and for fixed $f_1, \ldots, f_k$ the scalar product is zero unless the injection uses all edges $f_1, \ldots, f_k$, which yields

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

where $N_G(f_1, \ldots, f_k) := |f : V \to [n] \text{ inj. uses edges } f_1 \cdots f_k|$. We now claim that

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

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

Let us split the k-th Hoeffding term further. Consider the number $\alpha(f_1, \ldots, f_k)$ of vertices that are used in the graph with 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, $\alpha(f_1,\dots,f_k) \in \{\alpha_k,\dots,2k\wedge|V|\}$. This gives the decomposition

$$T_k = \sum_{\alpha=\alpha_k}^{\min(|V|,2k)} T_{k,\alpha},$$

and

$$\widetilde{T}_G \coloneqq T_G - \mu_{n,p}(T_G) - f_1 = \sum_{k=2}^{|E|} \sum_{\alpha=\alpha_k}^{\min(|V|,2k)} T_{k,\alpha}. \tag{2.31}$$

Using the $L^q(\mu_{n,p})$ estimates from Proposition 1.2 gives for all $q \geq 2$

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

Let $c(n,p) \coloneqq (2|E|n^{|V|-2}p^{|E|-1}\sqrt{\binom{n}{2}p(1-p)})$. The L^q estimate given above can be used to show the multilevel concentration inequality (as in the proof of Theorem 1.1)

$$\mu_{n,p}\left(c(n,p)^{-1}|\widetilde{T}_G| \geq t\right) \leq e^2 \exp\left(-\frac{1}{C\sigma^2(p)} \min_{k=2,\dots,|E|} \min_{\alpha=\alpha_k,\dots,\min(2k,|V|)} t^{2/k} h_{n,k,\alpha}^{2/k}\right), \tag{2.32}$$

with $h_{n,k,\alpha} \coloneqq \left(\frac{c(n,p)}{(p(1-p))^{k/2}p^{|E|-k}\|A^{(k,\alpha)}\|_2}\right)$, and it remains to prove that

$$\min_{k=2,\dots,|E|} \min_{\alpha=\alpha_k,\dots,2k\wedge|V|} c(n,p)/(p(1-p)p^{|E|-k}\|A^{(k,\alpha)}\|_2) \to \infty.$$

We can estimate $\|A^{(k,\alpha)}\|_2$ from above as follows. If there are α vertices used by the edges $f_1,\dots,f_k$, there are at most $n^{|V|-\alpha}$ ways to have an injection of V into $[n]$ using the edges $f_1,\dots,f_k$, and there are at most n^α such combinations of $f_1,\dots,f_k$, and thus

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

and for $k \geq 2$ this gives

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

which is equivalent to $\min_{k=2,\dots,|E|}\min_{\alpha\in\{\alpha_k,\dots,2k\wedge|V|\}} np^{(k-1)/(\alpha-2)} \to \infty$, i.e.

$$np^{\max_{k=2,\dots,|E|}\max_{H\subset G:|E(H)|=k}\frac{|E(H)|-1}{|V(H)|-2}} = np^{\widetilde{d}(G)} \to \infty. \tag{2.33}$$

$\square$

We may replace the condition on $\widetilde{d}(G)$ in Theorem 2.9 by a condition on the maximal degree $d(G)$, though with an additional factor 2 which may lead to weaker results in general.

Corollary 2.10. *Let $G=(V,E)$ be any simple, connected graph. If $p=p(n)$ satisfies $np^{2d(G)}/\log(1/p) \to \infty$, then*

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

Proof. This follows immediately from Theorem 2.9. Indeed, since $(x-1)/(y-2) \leq 2x/y$ for all integers x, y with $x = 2, y = 3$, $x = 3, y = 4$ and $y \geq x \geq 3$, we obtain $\widetilde{d}(G) \leq 2d(G)$ (for any graph apart from K_2), and thus $np^{\widetilde{d}(G)} \geq np^{2d(G)} \to \infty$. $\square$

Remark. Similar calculations (and a proof of a central limit theorem for subgraph counts under non-optimal conditions) have been done in [NW88], interpreting subgraph counts as incomplete U-statistics and using the Hoeffding decomposition to prove the CLT. However, [NW88, Theorem 3.1] does not seem to be quite correct, since for triangles it requires a normalization of the subgraph count by $(n-2)\sqrt{p(1-p)n(n-1)/2}$, which does not converge to a normal distribution in general. As can be seen from the decomposition, the correct normalization is $(n-2)p^2\sqrt{p(1-p)n(n-1)/2}$ (see also Figure 2 below). In our approach, we additionally provide a quantification, i.e. we show that $T_3 - \mu(T_3) - p^2(n-2)f_1 = \mathcal{O}_{\mu_{n,p}}(n^{3/2})$ with exponentially decaying tails.

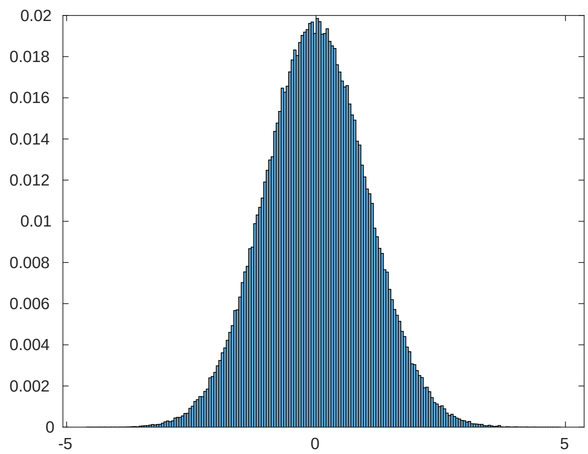

FIGURE 2. Histogram of $\frac{T_3 - \mu_{n,p}(T_3)}{(n-2)p^2\sqrt{p(1-p)n(n-1)/2}}$ in an Erdös-Renyi graph with $n = 1000$ and $p = 0.1$.

3. Proofs

In this section, we give the proofs of the results presented in Section 1. We start with Theorem 1.4. As indicated in Section 1, a key step is the application of an approximate tensorization result which goes back to K. Marton [Mar15] (see also [GSS18, Theorem 4.1]). For the reader's convenience, we state a version of it adapted to the situation considered in this article.

Theorem 3.1 (Marton). *Let μ be a spin system on $\mathcal{Y} = \mathcal{X}^{\mathcal{I}}$. If for some $\alpha_1, \alpha_2 > 0$*

$$\widetilde{\beta}(\mu) \geq \alpha_1 \qquad \textit{and} \qquad \|J\|_{2\to 2} \leq 1 - \alpha_2,$$

where J is a suitable interdependence matrix, then for any function $f : \mathcal{Y} \to \mathbb{R}_+$ vanishing outside of $\operatorname{supp}\mu$ *we have*

$$\mathrm{Ent}_\mu(f) \le \frac{1}{\alpha_1\alpha_2^2}\sum_{i\in\mathcal{I}}\int \mathrm{Ent}_{\mu(\cdot|\overline{x}_i)}(f(\overline{x}_i,\cdot))d\mu(x). \tag{3.1}$$

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 $i \notin S$.

We are now ready to prove our first result.

Proof of Theorem 1.4. Since μ satisfies the conditions of Theorem 3.1, we obtain that for any $f : \mathcal{Y} \to \mathbb{R}$ vanishing outside of $\Omega_0 := \operatorname{supp}\mu$

$$\mathrm{Ent}_\mu(f^2) \le \frac{1}{\alpha_1\alpha_2^2}\sum_{i\in\mathcal{I}}\int \mathrm{Ent}_{\mu(\cdot|\overline{y}_i)}(f^2(\overline{y}_i,\cdot))d\mu(y). \tag{3.2}$$

We can interpret this as a result on the probability space (Ω_0,μ), i.e. for any $f : \Omega_0 \to \mathbb{R}$ the inequality holds.

For any $i \in \mathcal{I}$, any $y \in \mathcal{Y}$ with $\mu(y) > 0$ the measure $\mu_n(\cdot \mid \overline{y}_i)$ is a measure on $\mathcal{X}$ with $\frac{1}{\min_{x\in\mathcal{X}}\mu(x|\overline{y}_i)} \le \frac{1}{\alpha_1}$, and thus by [BT06, Remark 6.6] we have

$$\mathrm{Ent}_{\mu(\cdot|\overline{y}_i)}(g^2) \le \frac{2\log(\alpha_1^{-1})}{\log(2)}\mathrm{Var}_{\mu(\cdot|\overline{y}_i)}(g),$$

which plugged into equation 3.2 leads to

$$\begin{aligned}\mathrm{Ent}_\mu(f^2) &\le 2\frac{\log(\alpha_1^{-1})}{\log(2)\alpha_1\alpha_2^2}\sum_{i\in\mathcal{I}}\int \mathrm{Var}_{\mu(\cdot|\overline{y}_i)}(f(\overline{y}_i,\cdot))d\mu(y)\\ &= 2\sigma^2\mathcal{E}_\mu(f,f).\end{aligned}$$

□

Proof of Theorem 1.1. First, note that since μ satisfies $\mathrm{LSI}(\sigma^2)$ with respect to $\mathfrak{d}$, by [GSS18, Proposition 2.4] we obtain for any $p \ge 2$

$$\|f-\mu(f)\|_p^2 \le \sigma^2 p\|\mathfrak{h}f\|_p^2. \tag{3.3}$$

Next we iterate (3.3) using $|\mathfrak{h}|\mathfrak{h}^{(d-1)}f|| \le |\mathfrak{h}^{(d)}f|$ (cf. [GSS18, Lemma 2.3]). This leads to

$$\|f-\mu(f)\|_p^2 \le \sum_{k=1}^{d-1}(\sigma^2 p)^k\|\mathfrak{h}^{(k)}f\|_2^2 + (\sigma^2 p)^d\|\mathfrak{h}^{(d)}f\|_p^2,$$

i. e. (1.9).

To prove the multilevel concentration bounds (1.10), we use methods outlined in [Ada06, Theorem 7], [AW15, Theorem 3.3] or [GSS18, Remark after Theorem 1.3]. To sketch the method in a slightly more general situation, assume that for any $p \ge 2$, we have for some constants $C_1, \dots, C_d \ge 0$

$$\|f-\mu(f)\|_p \le \sum_{k=1}^{d}(C_k p)^{k/2}$$

(which follows from (1.9) after taking roots). Write $N := |\{n : C_n > 0\}|$ and $r := \min\{k \in \{1, \ldots, d\} : C_k > 0\}$. By Chebyshev's inequality for any $p \geq 1$ we obtain

$$\mu(|f - \mu(f)| \geq e\|f - \mu(f)\|_p) \leq \exp(-p). \tag{3.4}$$

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 ∞, and assume that $\eta_f(t) \geq 2$, so that we can estimate

$$e\|f - \mu(f)\|_{\eta_f(t)} \leq e\sum_{k=1}^{d} \mathbb{1}_{C_k \neq 0} t = Net.$$

Applying equation (3.4) to $p = \eta_f(t)$ (if $p \geq 2$)

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

and combining it with the obvious estimate (in the case $p \leq 2$) gives

$$\mu(|f - \mu(f)| \geq (Ne)t) \leq e^2 \exp(-\eta_f(t)).$$

To remove the Ne factor, it is easiest to rescale the function by Ne and use the estimate $\eta_{(Ne)f}(t) \geq \frac{\eta_f(t)}{(Ne)^{2/r}}$. Thus, we have

$$\mu(|f - \mu(f)| \geq t) \leq e^2 \exp\left(-\frac{1}{(Ne)^{2/r}}\eta_f(t)\right).$$

□

To prove Theorem 1.2, let us introduce another notation. For any $l_1, \ldots, l_s \in \mathcal{I}$ 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 $A^{2=j,3=i}_{kl} = A_{kjil}$. 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 1.2. To see that $\mu(f_{d,A}) = 0$ fix $i_1, \ldots, i_d$ and let $P \in \mathcal{P}(\{i_1, \ldots, i_d\})$ be arbitrary. 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 $g_{\widetilde{P}} = \mu(g_P)$. As a consequence, $\mu(f_{d,A}) = 0$.

For any $l \in \mathcal{J}$ write T_l for the formal operator that replaces x_l by $\hat{x}_l$. We shall make use of the following inequality.

$$\begin{aligned}
\mathfrak{h}_l(f_{d,A}) &= \sup_{x_l,\hat{x}_l}\left|\sum_{I=(i_1,\ldots,i_d)} A_I \sum_{P\in\mathcal{P}(I)} (-1)^{M(P)}\left(g_P(x_I)) - g_P(T_l(x_I))\right)\right| \\
&= \sup_{x_l,\hat{x}_l}\left|\sum_{k=1}^{d}\sum_{I=(i_1,\ldots,i_{d-1})} A_I^{k=l} \sum_{\substack{P\in\mathcal{P}(I\cup\{l\})\\ \{l\}\in P}} (-1)^{M(P)}\left(g_P(x_I, x_l) - g_P(x_I, \hat{x}_l)\right)\right| \\
&= \sup_{x_l,\hat{x}_l}\left|(f(x_l) - f(\hat{x}_l))\sum_{k=1}^{d}\sum_{I=(i_1,\ldots,i_{d-1})} A_I^{k=l} \sum_{P\in\mathcal{P}(I)} (-1)^{M(P)} g_p(x_I)\right|
\end{aligned}$$

$$\leq c\left|\sum_{k=1}^{d}\sum_{I=(i_1,\dots,i_{d-1})}A_I^{k=l}\sum_{P\in\mathcal{P}(I)}(-1)^{M(P)}g_P(x_I)\right|$$
$$= c\left|\sum_{k=1}^{d}f_{d-1,A^{k=l}}\right|.$$

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

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

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

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

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

where $\widetilde{\mathfrak{h}}_l$ replaces $\sup_{x_l,\hat{x}_l}|f(x_l)-f(\hat{x}_l)|$ by 1. Clearly, since $\mathfrak{h}_lf_{d,A} \leq \widetilde{\mathfrak{h}}_lf_{d,A}$ 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

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

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 can be used and

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

Thus, in the d-th step we have

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

Taking the square root proves the claim. The multilevel concentration follows as in the proof of Theorem 1.1. □

References

[Ada06] Radosław Adamczak. "Moment inequalities for U-statistics". In: *Ann. Probab.* 34.6 (2006), pp. 2288–2314. DOI: 10.1214/009117906000000476.

[AW15] Radosław Adamczak and Paweł Wolff. "Concentration inequalities for non-Lipschitz functions with bounded derivatives of higher order". In: *Probab. Theory Related Fields* 162.3-4 (2015), pp. 531–586. DOI: 10.1007/s00440-014-0579-3.

[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.

[Bil68] Patrick Billingsley. *Convergence of probability measures.* John Wiley & Sons, Inc., New York-London-Sydney, 1968, pp. xii+253. ISBN: 978-0-471-19745-4.

[BGS17] Sergey G. Bobkov, Friedrich Götze, and Holger Sambale. “Higher Order Concentration of Measure”. In: *arXiv preprint* (2017). arXiv: 1709.06838.

[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.

[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.

[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/aoap/1034968224.

[GS16] Friedrich Götze and Holger Sambale. “Second Order Concentration via Logarithmic Sobolev Inequalities”. In: *arXiv preprint* (2016). arXiv: 1605.08635.

[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.

[Lat06] Rafał Latała. “Estimates of moments and tails of Gaussian chaoses”. In: *Ann. Probab.* 34.6 (2006), pp. 2315–2331. DOI: 10.1214/009117906000000421.

[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.

[Mar15] Katalin Marton. “Logarithmic Sobolev inequalities in discrete product spaces: a proof by a transportation cost distance”. In: *arXiv preprint* (2015). arXiv: 1507.02803.

[NW88] Krzysztof Nowicki and John C Wierman. “Subgraph counts in random graphs using incomplete U-statistics methods”. In: *Annals of Discrete Mathematics*. Vol. 38. Elsevier, 1988, pp. 299–310. DOI: 10.1016/0012-365X(88)90220-8.

[Ruc88] Andrzej Ruciński. “When are small subgraphs of a random graph normally distributed?” In: *Probab. Theory Related Fields* 78.1 (1988), pp. 1–10. DOI: 10.1007/BF00718031.

[Sin18] Arthur Sinulis. “Mixing times of Glauber dynamics via entropy methods”. In: *arXiv preprint* (2018). arXiv: 1804.04424.

[Tal96] Michel Talagrand. “A new look at independence”. In: *Ann. Probab.* 24.1 (1996), pp. 1–34. DOI: 10.1214/aop/1042644705.

[Wol13] Paweł Wolff. “On some Gaussian concentration inequality for non-Lipschitz functions”. In: *High dimensional probability VI*. Vol. 66. Progr. Probab. Birkhäuser/Springer, Basel, 2013, pp. 103–110.

Holger Sambale, Faculty of Mathematics, Bielefeld University, Bielefeld, Germany
E-mail address: hsambale@math.uni-bielefeld.de

Arthur Sinulis, Faculty of Mathematics, Bielefeld University, Bielefeld, Germany
E-mail address: asinulis@math.uni-bielefeld.de