REGULARITY METHOD AND LARGE DEVIATION PRINCIPLES FOR THE ERDŐS–RÉNYI HYPERGRAPH

NICHOLAS A. COOK, AMIR DEMBO, and HUY TUAN PHAM

Abstract

We develop a quantitative large deviations theory for random hypergraphs, which rests on tensor decomposition and counting lemmas under a novel family of cut-type norms. As our main application, we obtain sharp asymptotics for joint upper and lower tails of homomorphism counts in the r-uniform Erdős–Rényi hypergraph for any fixed $r \geq 2$, generalizing and improving on previous results for the Erdős–Rényi graph ($r = 2$). The theory is sufficiently quantitative to allow the density of the hypergraph to vanish at a polynomial rate, and additionally yields tail asymptotics for other nonlinear functionals, such as induced homomorphism counts.

Contents

DUKE MATHEMATICAL JOURNAL
Vol. 173, No. 5, © 2024 DOI 10.1215/00127094-2023-0029
Received 1 February 2022. Revision received 7 April 2023.
2020 *Mathematics Subject Classification*. 05C65, 60F10, 15A69, 05C80.

1. Introduction

1.1. Overview

For a fixed integer $r \geq 2$ and (large) integer n, let $\mathcal{Q}_{n,r} = [0,1]^{\binom{[n]}{r}}$ denote the set of $[0,1]$-valued functions on r-sets $I = \{i_1, \ldots, i_r\} \subset [n] := \{1, \ldots, n\}$. We associate elements of $\mathcal{Q}_{n,r}$ with edge-weighted r-uniform hypergraphs over $[n]$ with edge weights $Q(I)$, $I \in \binom{[n]}{r}$. The set $\mathcal{Q}_{n,r}$ also parameterizes the collection of inhomogeneous Erdős–Rényi measures μ_Q over *unweighted* r-uniform hypergraphs (r-*graphs*), where for a random r-graph with distribution μ_Q, each r-set I is included as an edge independently with probability $Q(I)$. For the case that $Q \equiv p$ for some $p \in (0,1)$, we have that $\mu_Q =: \mu_p$ is the distribution of the Erdős–Rényi r-graph $\boldsymbol{G} = \boldsymbol{G}^{(r)}_{n,p}$.

Our aim is to establish precise estimates, at exponential scale, for probabilities of rare events for $\boldsymbol{G}$, and in particular to justify asymptotics (in the large n limit) of the form

$$\log \mu_p(\mathcal{E}) \sim -\inf\{\mathrm{I}_p(Q) : Q \in \mathcal{E}'\} \tag{1.1}$$

for general sets of hypergraphs $\mathcal{E}$ (viewed as subsets of the discrete cube $\{0,1\}^{\binom{[n]}{r}}$), where the infimum is taken over an appropriate "approximation" $\mathcal{E}'$ of $\mathcal{E}$ in the solid cube $\mathcal{Q}_{n,r}$. Here, $\mathrm{I}_p(Q) := \mathrm{D}(\mu_Q \| \mu_p)$ is the relative entropy of μ_Q with respect to μ_p (see (1.17)).

Of particular interest are tail estimates for the number of occurrences of a fixed r-graph H as a sub-hypergraph of $\boldsymbol{G}$, which have been the subject of intense activity in recent years, mainly for the case $r = 2$ (we review the literature in Section 1.3 below). Writing $\mathsf{V}(H)$, $\mathsf{E}(H)$ for the vertex and edge sets of an r-graph H, and $\mathsf{v}(H)$, $\mathsf{e}(H)$ for their respective cardinalities, we recall that the *homomorphism density* of H in a weighted hypergraph $Q \in \mathcal{Q}_{n,r}$ is

$$t(H,Q) = \frac{1}{n^{\mathsf{v}(H)}} \sum_{\phi:\mathsf{V}(H)\to[n]} \prod_{e\in\mathsf{E}(H)} Q\big(\phi(e)\big), \tag{1.2}$$

where we have extended Q symmetrically to a function on ordered r-tuples, taking value zero when the coordinates are not all distinct. For the case that $Q = A_G$ is the 0–1 adjacency tensor of an r-graph G, this is the probability that a uniform random mapping of the vertices of H into the vertices of G maps the edges of H onto edges of G. In this case, we abusively write $t(H,G) := t(H, A_G)$.

As an application of our main results—namely, the quantitative LDP of Theorem 3.1 (a consequence of a tensor *decomposition lemma* (Theorem 2.13)) and a *counting lemma* (Theorem 2.15)—we obtain the following instances of (1.1) for intersections of super/sublevel sets of functionals (1.2). For a sequence of r-graphs

$\underline{H} = (H_1, \dots, H_m)$ and $\underline{\delta} = (\delta_1, \dots, \delta_m) \in \mathbb{R}_+^m$, define the joint upper-tail rate and corresponding entropic optimization problem

$$\mathrm{UT}_{n,p}(\underline{H}, \underline{\delta}) := -\log \mathbb{P}\big(t(H_k, \boldsymbol{G}) \geq (1+\delta_k) p^{\mathsf{e}(H_k)}, 1 \leq k \leq m\big), \tag{1.3}$$

$$\Phi_{n,p}(\underline{H}, \underline{\delta}) := \inf_{Q \in \mathcal{Q}_{n,r}} \big\{\mathrm{I}_p(Q) : t(H_k, Q) \geq (1+\delta_k) p^{\mathsf{e}(H_k)}, 1 \leq k \leq m\big\}, \tag{1.4}$$

and for $\underline{\delta} \in (0,1)^m$ the analogous joint lower-tail quantities

$$\mathrm{LT}_{n,p}(\underline{H}, \underline{\delta}) := -\log \mathbb{P}\big(t(H_k, \boldsymbol{G}) \leq (1-\delta_k) p^{\mathsf{e}(H_k)}, 1 \leq k \leq m\big), \tag{1.5}$$

$$\Psi_{n,p}(\underline{H}, \delta) := \inf_{Q \in \mathcal{Q}_{n,r}} \big\{\mathrm{I}_p(Q) : t(H_k, Q) \leq (1-\delta_k) p^{\mathsf{e}(H)}, 1 \leq k \leq m\big\}. \tag{1.6}$$

The scaling by $p^{\mathsf{e}(H)}$ is natural as one checks that $\mathbb{E}\, t(H, \boldsymbol{G}) \sim p^{\mathsf{e}(H)}$ for the range of p considered below. For an r-graph H, we write $\Delta(H)$ for its max-degree—that is, the maximum over $v \in \mathsf{V}(H)$ of $\deg_H(v) = |\{e \in \mathsf{E}(H) : v \in e\}|$. In the following, we additionally refer to a hypergraph parameter $\Delta'(H)$ whose definition is deferred to (3.15), only noting here that it always lies in the range

$$\frac{1}{r}\big(\Delta(H)+1\big) \leq \Delta'(H) \leq \Delta(H) + 1 \tag{1.7}$$

with the lower bound attained (for instance) by stars, and the upper bound attained by cliques. For our conventions on asymptotic notation, see Section 1.6.

THEOREM 1.1

Fix r-graphs $H_1, \dots, H_m$. Let $\Delta_{\max} = \max_k \Delta(H_k)$ and $\Delta'_{\max} = \max_k \Delta'(H_k)$.

(a) *(Joint upper tail). If $n^{-1/\Delta'_{\max}} \ll p < 1$, then for any fixed $\delta_1, \dots, \delta_m > 0$,*

$$\mathrm{UT}_{n,p}(\underline{H}, \underline{\delta}) \geq \big(1 - o(1)\big)\Phi_{n,p}\big(\underline{H}, \underline{\delta} - o(1)\big) \tag{1.8}$$

and if $n^{-1/\Delta_{\max}} \ll p < 1$, then

$$\mathrm{UT}_{n,p}(\underline{H}, \underline{\delta}) \leq \big(1 + o(1)\big)\Phi_{n,p}\big(\underline{H}, \underline{\delta} + o(1)\big). \tag{1.9}$$

(b) *(Joint lower tail). If $n^{-1/\Delta'_{\max}} \log n \ll p < 1$, then for any fixed $\delta_1, \dots, \delta_m \in (0,1)$,*

$$\mathrm{LT}_{n,p}(\underline{H}, \underline{\delta}) = \big(1 + o(1)\big)\Psi_{n,p}\big(\underline{H}, \underline{\delta} + o(1)\big). \tag{1.10}$$

Remark 1.2

For the proofs of (1.8) and (1.9), we may assume that $p \leq \min_k (1+\delta_k)^{-1/\mathsf{e}(H_k)}$, as otherwise the bounds hold vacuously. In Section 7.3, we establish (1.9) by an alternative argument (similar to the one for the upper bound in (1.10)) which yields a

wider range of p for certain graphs; we refrain from pursuing the widest range of p that can be obtained by our arguments under various assumptions on H. We remark that in the graphs setting ($r = 2$), where asymptotic formulas for $\Phi_{n,p}(\underline{H}, \underline{\delta})$ have been established (see [9], [10]), the upper bound (1.9) is easily obtained by computing the probability of specific events that saturate the upper tail. However, in the general r-graph setting such formulas have only been obtained in a few cases (see Corollary 3.2).

Remark 1.3
A similar result holds for mixed upper and lower tails for $p \gg n^{-1/\Delta'_{\max}} \log n$. However, when $p = o(1)$ the answer is just the lower-tail problem (1.5) for the subcollection of the H_k in the lower tail, together with any in the upper tail for which $\Delta(H_k) = 1$. This is due to the different speeds for lower- versus upper-tail deviations (of order $n^r p$ vs. $n^r p^{\Delta(H)}$). For similar reasons, it turns out that for $p = o(1)$ the right-hand sides of (1.8) and (1.9) are asymptotically equal to $\Phi_{n,p}(\pi(\underline{H}), \pi(\underline{\delta}))$, where π denotes restriction to the entries k for which $\Delta(H_k) = \Delta_{\min} = \min_\ell \Delta(H_\ell)$—see [10] for the case $r = 2$ (the idea is the same for general r) or the proof of Proposition 9.1. Consequently, the assumption for (1.8) can be relaxed to $1 > p \gg n^{-1/\Delta'_\star}$ for $\Delta'_\star = \max\{\Delta'(H_k) : \Delta(H_k) = \Delta_{\min}\}$.

Remark 1.4
By straightforward modifications of our arguments, Theorem 1.1 extends to random r-graphs drawn from inhomogeneous Erdős–Rényi measures μ_P, provided that $P(I) \in [cp, Cp]$ for all $I \in \binom{[n]}{r}$ and some fixed $0 < c < C < \infty$. (One replaces $\mathrm{I}_p(Q)$ with $\mathrm{D}(\mu_Q \| \mu_P)$ in (1.4) and (1.6).)

Various cases of Theorem 1.1 have been established before, mainly with $r = 2$ and/or $m = 1$, with some results holding in a wider range of p; we review the literature in Section 1.3 below. We note that treating joint tail events ($m \geq 2$) is important for applications to the analysis of general exponential random graph models, a class of Gibbs distributions on graphs that is widely applied in the social sciences literature (see [19], [20], [32], [40]).

Our aim is more general than joint tail estimates: in this work we initiate a quantitative large deviations theory for random hypergraphs. In particular, in Theorem 3.1 we establish versions of the approximation (1.1) for general sets $\mathcal{E}$ at (large) fixed n, which amount to quantitative large deviation principles (LDPs) for the Erdős–Rényi measure on r-graphs, extending the qualitative LDP of Chatterjee and Varadhan [21] for the case $r = 2$ and fixed p. The approximating sets $\mathcal{E}'$ are defined under a new family of tensor norms $\|\cdot\|^*_{\square}$ that generalize the matrix cut norm. The main technical

ingredient for establishing Theorem 3.1 is a *decomposition lemma* (Theorem 2.13) for sparse tensors that generalizes the classic Frieze–Kannan decomposition for matrices (see [42]). The role of the decomposition lemma is analogous to that of Szemerédi's regularity lemma in [21]. Combining with an accompanying *sparse counting lemma* (Theorem 2.15)—a deterministic result establishing sharp Lipschitz control on the functionals $t(H,\cdot)$ under the B^*-norms—we obtain the upper- and lower-tail bounds for homomorphism counts as contractions of the general LDPs.

We expect that our results could be applied or extended to other natural distributions on r-graphs. For instance, apart from inhomogeneous Erdős–Rényi r-graphs (see Remark 1.4 above) one may apply the results of this work to random regular hypergraphs, in a similar way to how large deviations results for the case $r = 2$ from [31] were applied to random regular graphs in [10] and [49].

The B^*-norms are the main innovation of this work. (The work [31] relied on the spectral norm, which is unavailable for tensors.) There are several novel features of these norms and the associated decomposition and counting lemmas. First, they are constructed to adapt to the level of sparsity under consideration. Second, as opposed to typical decomposition lemmas in extremal graph theory, our decomposition lemma attains a better quantitative bound that is crucial for obtaining Theorem 3.1, by (necessarily) excluding an exceptional set whose probability can be made arbitrarily small. Finally, both our tensor norm and decomposition lemma make explicit use of the Boolean nature of the test tensors in order to obtain the nearly optimal quantitative bound—in particular, in the case $r = 2$ we improve on the result from [31] for counts of general graphs H.

In extremal graph theory, the combination of decomposition lemmas (and closely related *regularity lemmas*) with counting lemmas is known as a *regularity method*, and our results for the B^*-norms thus comprise a new regularity method for sparse hypergraphs, which we expect will have applications outside of large deviations theory.

Within the context of large deviations, the regularity method approach is quite flexible, and we demonstrate this with an application to the upper tail for *induced* homomorphism counts in Theorem 10.1. We further obtain strong results for the lower tail of counts of Sidorenko hypergraphs in Theorem 10.2.

In Section 1.2 we review the connections between large deviations problems, graph limits, and the regularity method, highlighting a special case of one of our key technical results, the sparse counting lemma. In Section 1.3 we give an overview of previous works on upper and lower tails for random graphs, and in Section 1.4 we discuss the potential scope and limitations of the regularity method approach to quantitative LDPs.

1.2. Large deviation principles and the regularity method

On a conceptual level, the most important antecedent for our results is the seminal work of Chatterjee and Varadhan [21] establishing an LDP for the Erdős–Rényi graph (the case $r = 2$). Their result is a true LDP in the classical sense, in that it establishes asymptotics of the form (1.1) for subsets $\mathcal{E}$ of a fixed topological space $\mathcal{Q}$, where $\mathcal{E}'$ is an open/closed approximation of $\mathcal{E}$. It is perhaps unclear how such an LDP could be formulated in this context, as the Erdős–Rényi measures are on a *sequence* of spaces $\mathcal{Q}_{n,2}$ of growing dimension, but an appropriate setting is provided by the topological space of *graphons*, which is in some sense the completion of the collection of all finite graphs of all sizes under a topology induced by the cut norm. This is the appropriate topology for studying homomorphism densities $t(H,\cdot)$, as these extend to continuous functionals on graphon space—a consequence of the classic counting lemma. The key ingredient for the LDP is the compactness of graphon space, which is a consequence of Szemerédi's regularity lemma (in fact the Frieze–Kannan *weak regularity lemma* in [42] suffices for their purposes). Indeed, graphon theory gives a topological perspective on the classic regularity method in extremal graph theory, which is based on the regularity and counting lemmas.

We note that while large deviations theory was first formulated at the (in some sense "correct") level of generality of a topological theory by Varadhan [79] in the 1960s, the topological theory of dense graph limits was developed much more recently by Lovász, Szegedy, and coauthors (see [13]–[15], [63]). We refer to the books [18] and [62] for further background on graph limits and the regularity method.

Unfortunately, graphon theory is largely unsuitable for the study of sparse graphs, such as Erdős–Rényi graphs with $p = n^{-c}$ for any positive constant $c > 0$. In [17], Chatterjee poses the problem of developing a sparse graph limit theory that is powerful enough to prove upper-tail asymptotics for sparse Erdős–Rényi graphs. While there are by now several sparse graph limit theories (see, e.g., [11] and references therein), we do not know of any that are generally suitable for the study of large deviations.

The present work bypasses the development of an appropriate sparse (hyper)graph limit theory by instead developing a sparse hypergraph regularity method at finite n. As with sparse graph limits, existing sparse regularity tools are unsuitable for the study of upper-tail large deviations (we review the literature in Section 2). The main challenge in this context is *localization phenomena*: that the underlying mechanisms for upper-tail deviations in the sparse setting are the appearance of dense configurations of $o(n^2)$ edges, which are invisible to the cut-norm topology. Such localized structures are a general problem for the development of sparse extensions of the regularity and counting lemmas, and hence for a sparse graph limit theory. We remark that such localization phenomena do not occur in the

corresponding problem for lower-tail deviations. As such, asymptotics for extremes of the lower tail (the probability of containing no copy of a certain graph H) have been obtained previously in beautiful works around the KŁR conjecture and the hypergraph container method (see [4], [66], [71]; see also Remark 2.4).

In place of the cut norm, we introduce a family of tensor norms designed to detect localization phenomena. A key result is a (deterministic) sparse counting lemma giving optimal Lipschitz control on homomorphism counts of fixed r-graphs H in large sparse r-graphs G, which we expect could be useful for other extremal problems where localization plays an important role. We highlight here a special case of our sparse counting lemma for homomorphism counts of $K_4^{(3)}$, the complete 3-graph on 4 vertices (thus $K_4^{(3)}$ contains all 4 possible edges of size 3). Recall that the symmetric adjacency tensor for an r-graph G is denoted A_G. We say that H' is a proper sub-hypergraph of H if $\mathsf{V}(H') \subseteq \mathsf{V}(H)$ and $\mathsf{E}(H')$ is a strict subset of $\mathsf{E}(H)$.

THEOREM 1.5 (Sparse counting lemma for $K_4^{(3)}$-counts)
Let $p \in (0,1)$ be arbitrary, and let G_1, G_2 be two 3-graphs over the common vertex set $[n]$ such that

$$\max_{I,J,K\subseteq[n]^2} \Big| \sum_{(i,j,k)\in I^{(2,3)}\cap J^{(1,3)}\cap K^{(1,2)}} A_{G_1}(i,j,k) - A_{G_2}(i,j,k) \Big|$$
$$\le \varepsilon p \cdot \big(n^3 p^3 + np^2\big(|I| + |J| + |K|\big) + |I^{(2,3)} \cap J^{(1,3)} \cap K^{(1,2)}|\big) \quad (1.11)$$

for some $\varepsilon \in (0,1]$, where for $I \subset [n]^2$ and $a,b \in \{1,2,3\}$ we write $I^{(a,b)} := \{(i_1,i_2,i_3) : (i_a,i_b) \in I\} \subseteq [n]^3$. Assume further that

$$t(H, G_1) \le L p^{\mathsf{e}(H')} \quad (1.12)$$

for some $L \ge 1$ and all proper sub-hypergraphs $H' \subset K_4^{(3)}$. Then

$$\big|t(K_4^{(3)}, G_1) - t(K_4^{(3)}, G_2)\big| \lesssim \varepsilon L p^4. \quad (1.13)$$

The left-hand side of (1.11) is the maximal edge discrepancy between G_1 and G_2 over sets of the special form $I^{(2,3)} \cap J^{(1,3)} \cap K^{(1,2)}$, which play an analogous role to the cut sets $I \times J \subseteq [n]^2$ in the definition of the cut norm for 2-graphs. For homomorphism densities of general r-graphs H we consider edge discrepancies across structured sets with more general shapes, with carefully chosen (n,p)-dependent weights as on the right-hand side of (1.11), which will be crucial for obtaining accurate control when G_1, G_2 are sparse. The shapes of structured sets and the weights are summarized by a *weighted base system* $\mathbb{B}$, which leads to the definition of a norm $\|\cdot\|_{\mathbb{B}}^*$. In that setup, the bound (1.11) is equivalent up to constant factors to a bound of the form $\|A_{G_1} - A_{G_2}\|_{\mathbb{B}}^* \le \varepsilon p$ for a certain base system $\mathbb{B}$ (see Examples 2.7 and 2.9).

The general sparse counting lemma of Theorem 2.15 roughly states that for a given r-graph H, if G_1, G_2 are two (large) r-graphs such that

$$t(H', G_i) = O(p^{\mathsf{e}(H')}), \quad i = 1, 2, \tag{1.14}$$

for all proper sub-hypergraphs $H' \subset H$ (in particular G_1, G_2 are $O(p)$-sparse, by the case that H' is a single edge) and $\|A_{G_1} - A_{G_2}\|_{\mathbb{B}}^* \leq \varepsilon p$ for an appropriate choice of weighted base system $\mathbb{B}$ depending on H, then

$$\left|t(H, G_1) - t(H, G_2)\right| \lesssim \varepsilon p^{\mathsf{e}(H)}, \tag{1.15}$$

where the implicit constant depends only on H.

The full definition of the B^*-norms is a bit notationally involved (as is common in hypergraph regularity theory), so we first motivate them in Section 2 with a special instance for matrices. The key point is that the right-hand side of (1.15) can be made small compared to the typical value $\sim p^{\mathsf{e}(H)}$, even when $p = o(1)$ (the result is nonasymptotic so p may depend in an arbitrary way on n).

Sparse counting lemmas for the cut norm have been a subject of intense study ever since a sparse extension of Szemerédi's regularity lemma was observed by Kohayakawa [58] and Rödl. Existing sparse counting lemmas establish (1.15) under different hypotheses, generally assuming that G_1 and G_2 are both contained in a sparse pseudorandom "host" graph—effectively ruling out localization phenomena, which are treated as a nuisance—while only assuming that they are close in the cut metric, which is sensitive to differences in edge counts only at a macroscopic scale (over a constant proportion of the vertices). While such versions are effective for obtaining sparse Ramsey/Turán theorems or certain extreme cases of the lower tails, they are unsuitable for our purposes of controlling upper tails, which are *governed* by localization phenomena. Our assumption (1.14) is weaker than a pseudorandom host condition, while closeness under a B^*-norm is (necessarily) stronger, as these are designed to be sensitive to localization. We discuss these points further in Section 2.

The accuracy of the B^*-norms is only useful for large deviations if the space $\mathcal{Q}_{n,r}$ is sufficiently compact under these norms, in a quantitative (metric entropy) sense. Indeed, a typical approach to derive a large deviation upper bound is to first derive an upper bound on certain special sets, and combine them by constructing a covering of (most of) the space by these special sets. As encountered later in Theorem 6.1, by a straightforward consequence of the minimax theorem, one has a nonasymptotic large deviation upper bound taking the form of the right-hand side of (1.1) for the measure of convex sets $\mathcal{E}'$. Thus, one obtains large deviation upper bounds for more general sets $\mathcal{E}$ by covering them with convex sets and applying the union bound, leading to an error term given by the metric entropy of the set $\mathcal{E}$ (i.e., the logarithm of the covering number).

Suitable control on the metric entropy is established by the decomposition lemma (Theorem 2.13), which allows general sets $\mathcal{E}$ to be efficiently covered by small balls centered on "structured" tensors. Theorem 2.13 generalizes the Frieze–Kannan decomposition for matrices, and crucially provides more efficient decompositions after the (optional) removal of a small set of exceptional tensors.

As an example, in the context of $K_4^{(3)}$-counts as in Theorem 1.5 above, Theorem 2.13 implies that the set of all 3-graphs G over $[n]$ is covered by the εp-neighborhood (under the norm $\|\cdot\|_{\square}^*$) of a small collection of "structured" weighted 3-graphs, together with a set $\mathcal{E}$ of "exceptional" 3-graphs of measure $\mathbb{P}(\boldsymbol{G}_{n,p} \in \mathcal{E}) \le p^{Ln^3p^3}$. The structured weighted r-graphs have adjacency tensors Q that are linear combinations of at most $O(L\varepsilon^{-2}p^{-2})$ Boolean "test tensors" of the form $1_{(i,j,k)\in I^{(2,3)}\cap J^{(1,3)}\cap K^{(1,2)}}$, with notation as in Theorem 1.5. The parameter $L \ge 1$ is free to be chosen according to one's needs; note there is a trade-off between the measure of the exceptional set $\mathcal{E}$ and the size of the covering.

The decomposition lemma thus allows us to cover superlevel and sublevel sets for homomorphism densities $t(H,\cdot)$, by a small number of sets of diameter $O(\varepsilon p)$ in the appropriate $\square^*$-norm, together with an exceptional set whose measure can be made small compared to the large deviation rate. The counting lemma then shows that $t(H,\cdot)$ can only change by $O(\varepsilon p^{\mathsf{e}(H)})$ on these sets (recall that (1.11) is equivalent to such a bound) which allows us to justify the approximation of the upper and lower tails by (1.4) and (1.6), respectively.

1.3. Previous works

The past decade has seen several results of the form of Theorem 1.1 established for various ranges of sparsity p, mainly for the case $r = 2$ (the Erdős–Rényi graph) and $m = 1$, and often focusing only on the upper or lower tail. In the present work we aim for broader LDP-type statements as in (1.1), which can only be expected to hold in a proper subset of the range of p for which the asymptotics (1.8)–(1.10) are expected to hold—we discuss this point further in Section 1.4 below.

Many works have obtained asymptotics for $\mathrm{UT}_{n,p}(H,\delta)$ and $\mathrm{LT}_{n,p}(H,\delta)$ holding up to constants depending on δ. For the lower tail this is accomplished by Janson's inequality (see [51], [53]). For the "infamous" upper tail, following the works [52] and [57] obtaining upper and lower bounds matching up to a factor $\log(1/p)$, the sharp dependence on n and p was obtained in a wide range of p for triangles (see [16], [34]), general cliques (see [33]), cycles (see [70]), and stars (see [74]).

Following the LDP of [21] for fixed p and $r = 2$, the first results establishing sharp asymptotics for $\mathrm{UT}_{n,p}(H,\delta)$ allowing $p = n^{-c}$ took a rather different route from the one taken here, proceeding through a general study of Gibbs measures on the hypercube. This began with the influential work of Chatterjee and Dembo [19]

introducing a new *nonlinear large deviations* paradigm, further developed in [2], [3], [38], [39], and [80], with the focus of establishing sufficient conditions for validity of the naive mean-field approximation for the partition function, a problem of independent interest in statistical physics. Large deviation estimates were deduced through (lossy) approximation arguments, and hence these works could only permit a small sparsity exponent c.

The more direct approach to the large deviations problem through an appropriate finite-n regularity method was introduced by the first two authors in [31] for the case $r = 2$, where improved ranges for p were obtained by replacing the cut norm with the spectral norm. The sparse counting lemma step in that work was only sharp for cycle counts, which is ultimately due to the fact that these can be expressed as moments of the spectral distribution of the adjacency matrix. For the case of cycle counts, similar results (and superior in the case of triangles) were independently obtained by Augeri [2]. The method was further applied to edge eigenvalues of the adjacency matrix in [31], with a formula for the corresponding entropic optimization problem obtained in [8]; results on edge eigenvalues of sparser Erdős–Rényi graphs have more recently appeared in [5] and [7].

The lack of a spectral theory for tensors motivated the development of the B^*-norm regularity method, which is the main technical contribution of this work.

In [50], the upper-tail asymptotic (1.8)–(1.9) was extended to an essentially sharp range of p for the case ($r = 2$, $m = 1$, non-bipartite Δ-regular H), with an optimal result for the bipartite case subsequently obtained in [6]. (These works consider counts of *embeddings*, which only allow injective maps ϕ in (1.2); while the difference is negligible in the ranges of p considered here, embedding counts have significantly different behavior from homomorphism counts when $p \ll n^{-1/\Delta}$.) We comment further on the method of [6] and [50] in Section 1.4 below. Very recently (after the first version of this paper appeared on arXiv) the same method was further developed to obtain upper-tail asymptotics for *induced* homomorphism counts of $H = C_4$ in the Erdős–Rényi graph ($r = 2$) in an essentially optimal range of p (see [22]).

Upper tails for other random graph models besides the Erdős–Rényi distribution have been studied: $G(n, m)$ (uniformly random with n vertices and m edges) in [35]; random regular graphs in [10], [37], and [49]; and sparse inhomogeneous Erdős–Rényi graphs, such as stochastic block models in [10]. For the case of fixed p, extensions of the Chatterjee–Varadhan LDP to inhomogeneous Erdős–Rényi graphs have been established in [12], [47], and [67].

There are only a few works considering hypergraphs with $r \geq 3$. The upper-tail asymptotic (1.8)–(1.9) was established in [64] for the case of p fixed, $m = 1$ and H a *linear* hypergraph (see Example 3.8 below), and more recently in [61] for general H and $n^{-c(H)} \ll p \ll 1$ for sufficiently small $c(H) > 0$, using gen-

eral nonlinear large deviation tools from [38] (one checks that their proof allows $c(H) = 1/(6\mathsf{e}(H)) + o(1)$). Very recently, the lower-tail asymptotic (1.10) for the case $m = 1$ was established in an optimal range of sparsity in [60] by a beautiful entropy argument (as in [6] and [50] they consider embeddings rather than homomorphisms).

There is a parallel line of works establishing asymptotic formulas for the entropic optimization problems (1.4) and (1.6). For $r = 2$, $m = 1$, and fixed p, this was done in [21] and [64] for the upper tail in a certain region of the (p, δ)-plane. The latter work extended [21] to counts of linear hypergraphs in dense Erdős–Rényi hypergraphs, and further characterized the regime of (p, δ) for which the infimizer is the constant $Q \equiv p$ in this more general context. In [68], such a regime is provided for the variational problem corresponding to counts of general hypergraphs. For $r = 2$ and $1 \gg p \gg n^{-1/\Delta}$, an asymptotic formula was obtained for the upper tail for all $\delta > 0$ in [65] ($m = 1$, H a clique), [9] ($m = 1$, general H), and [10] (general m and H). In [81], Zhao obtains lower-tail formulas with p fixed or decaying as slowly as $n^{-c(H)}$ for a small $c(H) > 0$, and certain ranges of δ. For general r, $m = 1$, and $1 \gg p \gg n^{-1/\Delta}$, an asymptotic formula for the upper tail is obtained in [61] for the case H is a clique or the 3-graph depicted in Figure 1 (see Corollary 3.2).

Beyond establishing asymptotic formulas for (joint) upper and lower tails, there is the refined problem of describing the typical structure of $\boldsymbol{G}$ *conditioned* on the tail event. This has been addressed for p fixed in some cases in [21] and [64], and for the full range of $p = o(1)$ in [50] for the upper tail with $r = 2$, $m = 1$, and H a clique. More recently (after this paper was posted to the arXiv) the first two authors established the conditional structure of Erdős–Rényi graphs conditional on general joint upper tail events as in (1.3), with $p = o(1)$ allowed to decay at a certain (suboptimal) polynomial rates, by combining large deviations results of [31] and the present work with a stability analysis for solutions of the entropic optimization problem (1.4) established in [9] and [10]. In [32], the results on the conditional structure of Erdős–Rényi graphs were used to establish the *typical* structure of sparse exponential random graph models. We mention also the line of works [54]–[56], [69] on the related problem of determining the typical structure of dense random graphs with constrained edge and H counts for various choices of H.

1.4. Discussion

Our focus in this work is on the development of quantitative LDP-type statements as in (1.1) applying to general subsets of $\mathcal{Q}_{n,r}$ at large, fixed n, and to translate these to joint-tail asymptotics as in Theorem 1.1 using a sparse counting lemma. This approach has the advantage of being quite robust, applying to any functional enjoying a counting lemma under an appropriate B^*-norm—examples include nonmonotone

functionals such as *induced* homomorphism counts, as well as nonpolynomial functions such as the B^*-norms themselves (or compositions of these with affine maps, such as centering), which can be viewed as weighted generalizations of the max-cut functional. The method also applies *almost*[1] equally well to upper- and lower-tail events.

However, LDPs applying to general sets $\mathcal{E}$ can only be expected to hold in a limited range of sparsity. For instance, under the version of the B^*-norms that is needed to analyze clique counts, for which $\Delta(H) = \binom{v(H)-1}{r-1}$, the LDP only yields (1.8)–(1.9) for $n^{-1/(\Delta(H)+1)} \ll p < 1$, whereas the asymptotic should hold for all $p \gg n^{-r/\Delta(H)}$ (up to polylogarithmic factors). For general $\underline{H}$ we believe our method could be sharpened to yield the joint upper- and lower-tail asymptotics (1.8)–(1.10) in the range $1 > p \gg n^{-1/\Delta_{\max}}$; this would follow in particular from relaxing the condition (2.16) in the decomposition lemma by a factor p (see also Remark 2.3). While we can push further than $n^{-1/\Delta_{\max}}$ for certain $\underline{H}$ for which particularly efficient choices of B^*-norm suffice for an accurate counting lemma, in general we believe the upper and lower tails for $t(H, \boldsymbol{G})$ should require very different arguments when $p \ll n^{-1/\Delta(H)}$.

Thus, for certain sets $\mathcal{E}$, arguments establishing (1.1) in the optimal range of p will have to exploit special properties of $\mathcal{E}$ once p is below a certain threshold. For the case of superlevel sets for counts of a fixed regular graph H (the case of upper tails with $r = 2$, $m = 1$, and Δ-regular H in Theorem 1.1), this has been accomplished in the optimal sparsity range by a beautiful truncated moment method argument developed in [50] and further improved for the bipartite case in [6]. The general argument succeeds in covering the upper-tail event by events on which the discrete gradient of the subgraph-counting functional is essentially supported on a small set of edges that they call a "core," reducing the problem to the (quite technical) task of counting the possible locations of cores, which they accomplish by exploiting the special structure of subgraph-counting polynomials. In [50], they also apply their general method to the upper tail of counts of k-term arithmetic progressions in sparse subsets of $[n]$. While the method is simplest for upper tails of polynomials with nonnegative coefficients, it extends to certain nonmonotone polynomials including induced subgraph

[1]We say "almost" as there is a technical issue in applying the counting lemma for lower-tail estimates, stemming from the necessity of the crude upper bound (1.14) for counts of subgraphs. For upper tails this can be enforced by arguing inductively over the number of edges in $\boldsymbol{H}$, so that we can restrict to the high-probability event that such a bound holds for all smaller graphs. However, for the lower tail there is the issue that the bad event that (1.14) fails is of upper-tail type and hence is much larger than the event we want to estimate. For the proof of (1.10) we get around this by using the Fortuin–Kasteleyn–Ginibre (FKG) inequality to restrict to the event that (1.14) holds. This relies on monotonicity of homomorphism counts, which we do not have for induced homomorphism counts, and hence we do not have a result on the lower tail for the latter. We hope that an alternative argument for restriction to (1.14) can be found that avoids the use of monotonicity.

counts (see [50, Theorem 9.1], which is proved in the recent work [22] treating the upper tail for induced C_4-counts).

For lower tails, a beautiful entropic method was recently introduced in [60], where they obtain the asymptotic (1.10) (for embedding counts rather than homomorphism counts) in the optimal sparsity range. This approach makes use of the monotonicity of sublevel sets for embedding counts.

Finally, we note that whereas (1.8)–(1.10) are obtained by the sparse regularity method, we obtain the upper bound (1.9) for joint upper tails in the range $p \gg n^{-1/\Delta_{\max}}$ by applying a careful tilting argument, using the Efron–Stein inequality to derive concentration for homomorphism counts (as well as induced homomorphism counts) of a random tensor sampled from sparse product measures. In Section 7.3 we give an alternative argument, more along the lines of the proof of (1.8) and (1.10) and holding in a different range of p, which may be better or worse depending on $\underline{H}$.

1.5. Organization

In Section 2 we discuss previous extensions of the regularity method for sparse graphs, introduce the B^* tensor norms (first in the matrix case), and state our general decomposition and counting lemmas. Section 3 contains our general quantitative LDPs and some corollaries of Theorem 1.1 obtained by combining with earlier works on the upper-tail optimization problem $\Phi_{n,p}(H,\delta)$. Sections 4–6 contain the proofs of the counting lemma (Theorem 2.15), decomposition lemma (Theorem 2.13), and quantitative LDPs (Theorem 3.1). For the proof of Theorem 1.1, we establish (1.8) in Section 7, (1.10) in Section 8, and (1.9) in Section 9. In Section 10 we give extensions of Theorem 1.1 to induced homomorphism counts and the lower tail for counts of Sidorenko hypergraphs.

1.6. Notational conventions

We use C, c, c', and so on to denote constants that may change from line to line, understood to be absolute if no dependence on parameters (such as r) is indicated. For a (set of) parameter(s) P, we write $C(P)$ for a constant depending only on P.

(Standard) asymptotic notation

For quantities f, g depending on other parameters such as n or H, we write $f = O(g)$, $f \lesssim g$, and $g \gtrsim f$ to mean $|f| \le Cg$, and $f = \Theta(g)$ to mean $f \lesssim g \lesssim f$. We indicate dependence of the implied constant on parameters P by writing, for example, $f = O_P(g)$, $f \lesssim_P g$. Notation $o(\cdot)$, $\omega(\cdot)$e $\gg$, $\ll$ is with respect to the limit $n \to \infty$, with $f = o(g)$, $g = \omega(f)$, $f \ll g$, and $g \gg f$ being synonymous to the statement $f/g \to 0$.

Tensors

Throughout, we consider r fixed independently of n. Denote by $\mathcal{Z}_{n,r}$ the set of order-r tensors of size n (r-*tensors*), which we view as mappings $Z : [n]^r \to \mathbb{R}$. We equip $\mathcal{Z}_{n,r}$ with the usual ℓ_p-norms $\|Z\|_p^p = \sum_{i_1,\dots,i_r \in [n]} |Z(i_1,\dots,i_r)|^p$. The Euclidean inner product on $\mathcal{Z}_{n,r}$ for any r (including $\mathcal{Z}_{n,1} \cong \mathbb{R}^n$) is denoted $\langle \cdot, \cdot \rangle$. The orthogonal projection to a subspace W is denoted P_W. For a set $\mathcal{E} \subset \mathcal{Z}_{n,r}$, we write $\mathrm{hull}(\mathcal{E})$ for its convex hull.

An r-tensor is symmetric if it is invariant under permutation of the arguments. We write $\mathcal{S}_{n,r} \subset \mathcal{Z}_{n,r}$ for the set of symmetric r-tensors supported on entries with r distinct coordinates, $\mathcal{A}_{n,r} \subset \mathcal{S}_{n,r}$ for the subset of Boolean tensors, which are naturally associated to r-graphs, and $\mathcal{Q}_{n,r} := \mathrm{hull}(\mathcal{A}_{n,r})$, that is, the set of $Q \in \mathcal{S}_{n,r}$ with all entries lying in $[0,1]$. For $S \in \mathcal{S}_{n,r}$, we often abusively view its argument as an unordered set, writing $S(I) := S(i_1,\dots,i_r)$ for $I = \{i_1,\dots,i_r\}$.

Recall the distributions μ_Q introduced at the start of Section 1.1, which we view as measures on $\mathcal{A}_{n,r}$. We generally deal with random hypergraphs through their adjacency tensors in $\mathcal{A}_{n,r}$. Unless otherwise stated, $\mathbb{P}$ is a probability measure under which $\boldsymbol{A}$ has distribution μ_p, so that $\boldsymbol{A}$ is the adjacency matrix for the Erdős–Rényi hypergraph $\boldsymbol{G} = \boldsymbol{G}_{n,p}^{(r)}$, and $\mathbb{E}$ is the associated expectation. For $Q \in \mathcal{Q}_{n,r}$, we write $\mathbb{P}_Q, \mathbb{E}_Q$ for probability and expectation under which $\boldsymbol{A}$ has the distribution μ_Q. The relative entropy between the Bernoulli(p) and the Bernoulli(x) measures on $\{0,1\}$ is denoted

$$\mathrm{I}_p(x) := \mathrm{D}(\mu_x \| \mu_p) = x \log \frac{x}{p} + (1-x) \log \frac{1-x}{1-p}, \quad x \in [0,1] \tag{1.16}$$

(extended continuously from $(0,1)$ to $[0,1]$). With some abuse we use the same notation for the relative entropy of μ_Q with respect to μ_p on $\mathcal{A}_{n,r}$, defining

$$\mathrm{I}_p : \mathcal{Q}_{n,r} \to [0,\infty), \qquad \mathrm{I}_p(Q) = \sum_{1 \le i_1 < \cdots < i_r \le n} \mathrm{I}_p\big(Q(i_1,\dots,i_r)\big). \tag{1.17}$$

Note that $\mathbb{E}\boldsymbol{A} = p J_{n,r}$ for the adjacency tensor $J_{n,r}$ of the complete r-graph on n vertices. That is, $J_{n,r}(i_1,\dots,i_r) = 1$ if the indices are all distinct and zero otherwise.

Hypergraphs

All r-graphs are assumed to be finite and simple (i.e., with edge sets having no repeated elements). We often refer to r-uniform hypergraphs, sub-hypergraphs, and hyperedges simply as r-graphs, subgraphs, and edges, respectively. For hypergraphs $H = (\mathsf{V}, \mathsf{E})$ and $H' = (\mathsf{V}', \mathsf{E}')$, we say $H' \subseteq H$ if $\mathsf{V}' \subseteq \mathsf{V}$ and $\mathsf{E}' \subseteq \mathsf{E}$, and $H' \subset H$ if $\mathsf{V}' \subseteq \mathsf{V}$ and $\mathsf{E}' \subset \mathsf{E}$. We write $\mathsf{v}(H) := |\mathsf{V}(H)|$, $\mathsf{e}(H) := |\mathsf{E}(H)|$, and $\Delta(H)$ for the maximum degree of H, that is, the maximum number of edges sharing a common vertex $v \in \mathsf{V}(H)$. For $U \subset \mathsf{V}(H)$, we write

$$\partial^H U := \big\{e \in \mathsf{E}(H) : e \neq U, e \cap U \neq \emptyset\big\}, \qquad d^H(U) := |\partial^H U| \tag{1.18}$$

for the edge boundary of U and its cardinality, respectively (excluding U itself when it is an edge). For $\mathsf{f} \subset U \subseteq \mathsf{V}$, we denote the f-*dominated* boundary and degree of U by

$$\begin{aligned} \partial^H_{\mathsf{f}} U &:= \big\{e' \in \mathsf{E}(H) : \emptyset \neq e' \cap U \subseteq \mathsf{f}\big\} = \partial^H U \setminus \partial^H(U \setminus \mathsf{f}), \\ d^H_{\mathsf{f}}(U) &:= |\partial^H_{\mathsf{f}} U|. \end{aligned} \tag{1.19}$$

This is a subset of $\partial^H U$ consisting of edges whose overlap with U is contained in f. We additionally set $\partial^H_\emptyset U := \emptyset$, $d^H_\emptyset(U) := 0$. We will usually drop the superscript H from all notation, but in some places there will be more than one hypergraph in play and it will be necessary to clarify.

Homomorphism counts

As in several previous works on upper tails (e.g., [9], [19], [21]), we count subgraphs in the sense of hypergraph homomorphisms. Recall that a homomorphism between r-graphs H and G is a mapping $\phi : \mathsf{V}(H) \to \mathsf{V}(G)$ such that the image of every edge in H is an edge in G. We do not require that ϕ be injective—in particular, distinct edges of H may be mapped to a common edge in G. We write $\hom(H, G)$ for the number of homomorphisms from H to G, so that $t(H, G)$ from (1.2) is $n^{-\mathsf{v}(H)} \hom(H, G)$. We extend this to a function on $\mathscr{S}_{n,r}$ as

$$\hom(H, S) := \sum_{\phi : \mathsf{V}(H) \to [n]} \prod_{e \in \mathsf{E}(H)} S\big(\phi(e)\big) \tag{1.20}$$

so that for a graph G over $[n]$ with adjacency tensor A_G we have $\hom(H, G) = \hom(H, A_G)$. Here, with slight abuse we interpret $S(\phi(e))$ for $e = \{v_1, \dots, v_r\}$ as $S(\phi(v_1), \dots, \phi(v_r))$ when ϕ is injective on e and zero otherwise. We additionally denote the normalized quantities

$$t(H, S) := \frac{\hom(H, S)}{n^{\mathsf{v}(H)}}, \qquad t_p(H, S) := t(H, S/p) = \frac{\hom(H, S)}{n^{\mathsf{v}(H)} p^{\mathsf{e}(H)}}, \tag{1.21}$$

often writing $t(H, G) := t(H, A_G)$ and $t_p(H, G) := t_p(H, A_G)$.

2. Novel cut-type norms and a sparse regularity method

Our general approach reduces the problem of large deviations for nonlinear functionals of Erdős–Rényi hypergraphs to the development of a sparse hypergraph regularity method under appropriate extensions of the cut norm. This is a problem of general interest in extremal graph theory that goes beyond applications to large deviations, and there is already a large body of literature on sparse extensions of the regularity

method. In this section we begin with a brief overview of such results and explain why their assumptions make them unsuitable for our purposes. Then we discuss a special instance of the norms and decompositions in the matrix setting, in order to motivate the more complicated statements for general hypergraphs (deferred to Sections 2.3–2.4).

2.1. Previous work on sparse regularity

Much of the literature on sparse regularity is with an eye toward sparse extensions of classical Turán-type theorems. These show that sufficiently dense subsets G of a large set Γ are guaranteed to contain some small structure—specifically, a set from a distinguished class $\mathcal{S} \subset \binom{\Gamma}{k}$ of k-sets for some fixed k. For instance, if Γ is the edge set of the complete graph K_n on n-vertices, and $\mathcal{S}$ is the set of $\binom{r+1}{2}$-tuples of edges forming a copy of K_{r+1}, then Turán's theorem guarantees that $G \subset \Gamma$ contains some element of $\mathcal{S}$ when $|G|/|\Gamma|$ exceeds $1 - \frac{1}{r}$ (see [78]). A second example is Szemerédi's theorem in [77], where $\Gamma = [n]$, $\mathcal{S}$ is the collection of k-term arithmetic progressions, and G must contain some $S \in \mathcal{S}$ if $|G| \geq \delta|\Gamma|$ for any fixed positive δ and n sufficiently large.

Sparse Turán-type theorems establish the same statements when the "host set" Γ is instead taken to be a sparse pseudorandom subset of the host set Γ_0 from the corresponding classical theorem. An example is the Green–Tao theorem establishing existence of arithmetic progressions of arbitrary length in the primes [48, Theorem 1.2], which proceeded through a "relative Szemerédi theorem" for a certain set Γ of almost-primes that is a sparse pseudorandom subset of $\Gamma_0 = [n]$ (see [48]). In the realm of graph theory, Turán's theorem (and more generally, the Erdős–Stone theorem and Simonovits's stability theorem) has been transferred to host graphs Γ such as sparse Erdős–Rényi graphs (see [27], [72]; see also Remark 2.4 below for a discussion of related results) and graphs satisfying certain pseudorandomness conditions (see [25]).

These results can be proved by mimicking proofs of corresponding results for the dense setting, for instance, via sparse versions of the hypergraph removal lemma, which in turn are obtained from sparse extensions of hypergraph regularity and counting lemmas. Let us briefly recall these in the graph setting (2-uniform hypergraphs). Recall the normalized matrix cut norm

$$\|M\|_{\square} = \frac{1}{n^2} \max_{I,J \subseteq [n]} \left| \langle M, \mathbf{1}_I \otimes \mathbf{1}_J \rangle \right|, \quad M \in \mathbb{R}^{n \times n}. \tag{2.1}$$

Here, $\mathbf{1}_I \otimes \mathbf{1}_J$ is the rank-1 matrix $\mathbf{1}_I \mathbf{1}_J^{\mathsf{T}}$, and $\langle \cdot, \cdot \rangle$ is the Euclidean (Hilbert–Schmidt) inner product on $\mathbb{R}^{n \times n}$. This extends to a metric $d_{\square}$ on the set $\mathcal{G}_n$ of graphs over the vertex set $[n]$ as $d_{\square}(G_1, G_2) = \|A_{G_1} - A_{G_2}\|_{\square}$, with A_{G_i} the adjacency matrix for G_i. Thus, we trivially have $d_{\square}(G_1, G_2) \leq 1$, and a bound $d_{\square}(G_1, G_2) \leq \varepsilon < 1$

provides uniform control on the discrepancy between G_1, G_2 of edge counts in vertex subsets of $[n]$ of *macroscopic size*, that is, linear in n.

A result of Frieze and Kannan [42] (from which their weak regularity lemma is easily deduced) states that for any graph G, there is a decomposition of its adjacency matrix as

$$A_G = A_{\mathrm{str}} + A_{\mathrm{rand}}, \tag{2.2}$$

where the *structured* piece A_{str} is a linear combination of $O(1/\varepsilon^2)$ *cut matrices* $\mathbf{1}_{I_k} \otimes \mathbf{1}_{J_k}$, and the *pseudorandom* piece A_{rand} satisfies $\|A_{\mathrm{rand}}\|_\square \le \varepsilon$. The cut-norm counting lemma says that the homomorphism density functionals $t(H,\cdot)$ (recall (1.21)) are $O_H(1)$-Lipschitz in the cut metric. For graphs G_1, G_2 of density $p = o(1)$ we trivially have $d_\square(G_1, G_2) \lesssim p$, so for a sparse regularity lemma we seek a decomposition as in (2.2) with $\|A_{\mathrm{rand}}\|_\square \le \varepsilon p$. In general, such a decomposition requires a growing number of cut matrices, but straightforward modifications of the Frieze–Kannan argument yield sparse decomposition lemmas with $O_\varepsilon(1)$ cuts under additional "no dense spots" assumptions on G (see [23]).

One may similarly hope for a sparse counting lemma saying that $|t_p(H, G_1) - t_p(H, G_2)| \lesssim_H \varepsilon$ (recall the notation (1.21)) if $d_\square(G_1, G_2) \le \varepsilon p$, but this too is false without additional assumptions: consider, for instance, the case that G_1 and G_2 agree on all edges outside a set of vertices V_0 of size $\Theta(np)$, where G_1 is empty and G_2 is full. Since these only differ on $O(n^2p^2)$ edges we have $d_\square(G_1, G_2) \lesssim p^2 = o(p)$, whereas for triangle counts (say) we have $|t_p(K_3, G_1) - t_p(K_3, G_2)| \gtrsim 1$. However, in the applications to sparse Turán-type theorems described above, G_1 and G_2 are both contained in a pseudorandom (or truly random) host graph Γ, and sparse counting lemmas have been established under various "(pseudo)random container" assumptions (see [1], [25], [26], [28], [43]).

2.2. *A modified cut norm for sparse graphs*

Unfortunately, none of the sparse regularity or counting lemmas just described are useful for us, as the "no dense spots" and "pseudorandom container" hypotheses rule out the localization phenomena we are trying to detect. In [9] and [65], two phenomena are identified as the dominant mechanisms for upper-tail deviations of $t_p(H, \boldsymbol{G})$ in the Erdős–Rényi graph for $n^{-1/\Delta(H)} \ll p \ll 1$: the appearance of an almost-clique (of density close to 1) on $\Theta(np^{\Delta(H)/2})$ vertices,[2] or of an almost-complete bipartite graph on $J \times [n] \setminus J$ for $|J| = \Theta(np^{\Delta(H)})$. Both types of subgraphs contain $\Theta(n^2p^\Delta) = o(n^2)$ edges when $p = o(1)$ and are hence invisible to the cut norm; moreover, the cuts that correlate with $\boldsymbol{G}$ on these events, namely, $\mathbf{1}_{I_0} \otimes \mathbf{1}_{I_0}$ and $\mathbf{1}_{J_0} \otimes \mathbf{1}_{[n]\setminus J_0}$, have factors occurring at three separate scales.

[2]In fact, the almost-clique mechanism only contributes to large deviations when H is a regular graph.

Further localization phenomena have been described in the setting of regular graphs (see [10], [49]), and the possibilities are more numerous in the hypergraph setting (see [61]).

Our approach is to develop generalizations $\|\cdot\|_{\mathsf{B}}^*$ of the cut norm that are sensitive to localization phenomena at all scales. In the general hypergraph setting this is done in terms of a (user-specified) set system F over $[r]$, and the class of cut matrices is replaced by a class of *test tensors* T that are entrywise products of tensors $\tau_{\mathsf{f}}, \mathsf{f} \in \mathsf{F}$ varying only on coordinates in f. We defer the general definitions to Section 2.3 and discuss here a particular instance of these norms in the case of 2-graphs. (See also Theorem 1.5 and the discussion that follows it for an example for 3-graphs, stated there in terms of sets of edges rather than the functional formulation given here.)

Denote by $\mathcal{T} = \mathcal{T}_n$ the class of Bernoulli cut matrices $T = \mathbf{1}_I \otimes \mathbf{1}_J$ with $I, J \subseteq [n]$. Given a graph $H = (\mathsf{V}, \mathsf{E})$ of maximum degree Δ, we set a cutoff scale $n_0 := np^{\Delta-1}$ and for $T = \mathbf{1}_I \otimes \mathbf{1}_J \in \mathcal{T}$ denote

$$\|T\|_{\Delta,2} = \big(|I| \vee n_0\big)\big(|J| \vee n_0\big). \tag{2.3}$$

(This can be extended to a norm on $\mathbb{R}^{n\times n}$, but we only apply it to cut matrices.) Now for $M \in \mathbb{R}^{n\times n}$, let

$$\|M\|_{\Delta,2}^* = \sup_{T\in\mathcal{T}} \frac{|\langle M, T\rangle|}{\|T\|_{\Delta,2}} = \max_{I,J\subseteq[n]} \frac{|\,\mathbf{1}_I^{\mathsf{T}} M\, \mathbf{1}_J\,|}{(|I| \vee n_0)(|J| \vee n_0)}. \tag{2.4}$$

Note that $\|\cdot\|_{\Delta,2}$ and $\|\cdot\|_{\Delta,2}^*$ depend on p, but we suppress this from the notation. The $\|\cdot\|_{\Delta,2}^*$-norm specializes to the normalized cut norm (2.1) upon taking $p = 1$, but for smaller p the $\|\cdot\|_{\Delta,2}^*$-norm is sensitive to changes in density at smaller scales. The counting lemma of Theorem 2.15 specializes to this setting to say that if 2-graphs G_1, G_2 over $[n]$ satisfy $\|A_{G_1} - A_{G_2}\|_{\Delta,2}^* \le \varepsilon p$ and $t_p(H', G_1) \le L$ for every proper subgraph $H' \subset G_1$, then $|t_p(H, G_1) - t_p(H, G_2)| \lesssim_H L\varepsilon$. We note that the full version Theorem 2.15 generalizes this to multilinear homomorphism functionals, and Theorem 4.1 further extends to signed homomorphisms, which includes induced homomorphisms.

Note that some form of cutoff scale n_0 is necessary, as otherwise the norm would be too sensitive to changes on single entries. The specific choice $np^{\Delta-1}$ is motivated by the proof of the counting lemma, where it is a critical threshold for the influence of an endpoint of a single edge of G on $t(H, G)$. Indeed, by a telescoping decomposition based on the edges of H, one can express $t_p(H, G_1) - t_p(H, G_2)$ as a sum of terms indexed by edges $e = \{u, v\} \in \mathsf{E}(H)$ and embeddings $\psi : \mathsf{V}(H) \setminus \{u, v\} \to [n]$. Each term in the sum can be expressed in the form $\langle A_{G_1} - A_{G_2}, \mathbf{1}_I \otimes \mathbf{1}_J\rangle$, where I and J are the common neighbors of $\psi(\partial^H(u))$ and $\psi(\partial^H(v))$ (recall our notation (1.18)). In a random graph, $|I|$ and $|J|$ are typically of order $np^{\deg_H(u)-1}$ and $np^{\deg_H(v)-1}$,

respectively, which are at least $np^{\Delta-1}$. This is the motivation for the cutoff $n_0 = np^{\Delta-1}$ in the definition of $\|\cdot\|_{\Delta,2}$.

The following is a special case of our tensor decomposition lemma (Theorem 2.13). Recall that $\boldsymbol{A} = \boldsymbol{A}^{(2)}_{n,p}$ is the adjacency matrix for the Erdős–Rényi graph.

THEOREM 2.1 (Decomposition lemma, special case)
There exist absolute constants $C_0, c_0 > 0$ such that the following holds. Let $\kappa, \varepsilon > 0$, and assume that n and $p \in (n^{-2}, 1)$ are such that

$$np^{\Delta+1} \geq \frac{C_0 \log n}{\varepsilon^2 \log(1/p)}. \tag{2.5}$$

Then there exists a (possibly empty) exceptional set $\mathcal{E}_\star(\kappa, \varepsilon) \subseteq \mathcal{A}_{n,2}$ with $\mathbb{P}(\boldsymbol{A} \in \mathcal{E}_\star(\kappa, \varepsilon)) \leq p^{c_0 \kappa n^2}$ such that for each $A \in \mathcal{A}_{n,2} \setminus \mathcal{E}_\star(\kappa, \varepsilon)$ there is a decomposition

$$A = A_{\mathrm{str}} + A_{\mathrm{rand}}, \tag{2.6}$$

where

$$A_{\mathrm{str}} = pJ_{n,2} + \sum_{i=1}^{k} \alpha_i T_i \qquad \text{and} \qquad \|A_{\mathrm{rand}}\|^*_{\Delta,2} \leq \varepsilon p$$

for real numbers $\alpha_1, \ldots, \alpha_k$, and cut matrices $T_1, \ldots, T_k$ such that

$$\sum_{i=1}^{k} \|T_i\|_{\Delta,2} \leq \kappa \varepsilon^{-2} p^{-2} n^2. \tag{2.7}$$

Remark 2.2
Theorem 2.1 contains the Frieze–Kannan decomposition lemma (2.2) as a special case: taking $p = 1/2$ (say) makes $n^2\|\cdot\|^*_{\Delta,2}$ equivalent to the cut norm, and then taking κ to be a sufficiently large absolute constant makes $\mathcal{E}_\star(\kappa, \varepsilon) = \emptyset$; from the lower bound $\|T_i\|_{\Delta,2} \geq n_0^2 \gtrsim n^2$; it follows that $k = O(1/\varepsilon^2)$. However, the option to remove the exceptional set of tensors $\mathcal{E}_\star(\kappa, \varepsilon)$ is important for our application to upper tails, as taking a smaller value of κ reduces the complexity of the approximation A_{str}, effectively reducing the dimension of the space of tensors. We will set κ just large enough that $-\log \mathbb{P}(\mathcal{E}_\star(\kappa, \varepsilon))$ is above the large deviation rate (for instance, for the upper tail of $t(H, \boldsymbol{G})$ this is $\kappa = Kp^{\Delta(H)}$ for a sufficiently large constant K). A further key difference from the Frieze–Kannan decomposition lemma is that in (2.7) the complexity is measured in terms of the total size of the cut matrices.

Remark 2.3
We believe that the right-hand side of (2.7) can be improved to $\kappa\epsilon^{-2}p^{-1}n^2$, which

would allow us to replace the left-hand side of (2.5) with np^{Δ}. An assumption of $np^{\Delta} \gg 1$ would be essentially optimal: when $np^{\Delta} \lesssim 1$, the row-sums in submatrices of size $np^{\Delta-1} \times np^{\Delta-1}$ (the smallest scale controlled by the norm $\|\cdot\|_{\Delta,2}$) cease to concentrate, and we can no longer have uniform control on densities in such submatrices holding with high probability. Relaxing the left-hand side in (2.5) to np^{Δ}, and more generally saving a factor p in the analogous assumption (2.16) for the general decomposition lemma, would immediately imply that Theorem 1.1 holds with $\Delta'_{\max}$ replaced by $\Delta_{\max}$ in all cases. (For more precise details on the improvement in the case $r = 2$ as well as in the general case, we refer to Remark 3.10.)

In the standard way one can deduce a weak regularity lemma-type statement in terms of the partition of $[n]$ generated by the factors of the test tensors T_i, but this is not needed for our applications.

Theorem 2.1 takes the typical form of a decomposition lemma from graph theory and additive combinatorics, in that the summands in the expansion of the structured piece are controlled in some norm $\|\cdot\|$, while the pseudorandom piece is small in the dual norm $\|\cdot\|^*$, a general perspective that was explored by Gowers in [46]. Another common form of decomposition lemma obtains finer control on the pseudorandom piece, making $\|A_{\text{rand}}\|^*$ small relative to the "complexity" k of the structured piece, by separating out a further piece A_{small} that is small in another norm such as ℓ_2 (the original regularity lemma of Szemerédi is of this type). This comes at the cost of a much larger value of k than in weak regularity lemmas, and we will not need such control.

Before moving on to the general definition of the B^*-norms for r-tensors, let us mention one other case of the norms when $r = 2$, which is to take the smaller class $\mathcal{T} = \{\mathbf{1}_I \otimes \mathbf{1} : I \subseteq [n]\}$, with

$$\|T\|_{\Delta,1} = n \cdot \big(|I| \vee n_0\big) \quad \text{for } T = \mathbf{1}_I \otimes \mathbf{1} \tag{2.8}$$

and

$$\|M\|^*_{\Delta,1} = \sup_{T \in \mathcal{T}} \frac{|\langle M, T\rangle|}{\|T\|_{\Delta,1}} = \max_{I \subseteq [n]} \frac{|\,\mathbf{1}_I^{\mathsf{T}} M\, \mathbf{1}\,|}{n(|I| \vee n_0)}. \tag{2.9}$$

Alternatively, letting $d_M(i) := \frac{1}{n}\sum_{j=1}^n M(i,j)$ be the vector of normalized row sums for M,

$$\|M\|^*_{\Delta,1} = \max_{I \subseteq [n]} \frac{\langle d_M, \mathbf{1}_I\rangle}{|I| \vee n_0}.$$

This norm turns out to be effective for studying star homomorphism densities $t(K_{1,\Delta}, G)$, which is perhaps unsurprising since star homomorphism counts are determined by the degree sequence of G—indeed, we have $t(K_{1,\Delta}, G) = \frac{1}{n}\|d_{A_G}\|^{\Delta}_{\ell_\Delta}$. In

this case, our general decomposition lemma is approximating the degree sequence d_{A_G} of a graph by a short weighted combination of indicators $\mathbf{1}_{I_k}$, which can be done more efficiently than approximating the whole matrix A_G in the norm $\|\cdot\|^*_{\Delta,2}$. As a result, Theorem 1.1 gives tail asymptotics for star homomorphism counts in a wider range of p than for, say, clique counts. The form of (2.9) as compared to (2.4) illustrates a key point of the B^*-norms defined below: that to have an effective counting lemma for H-counts, we only need to use test tensors that are nonconstant on coordinates corresponding to vertices where edges overlap.

Remark 2.4
It would be remiss to not mention work on the KŁR conjecture on an embedding lemma for subgraphs of the sparse Erdős–Rényi graph $\boldsymbol{G}^{(2)}_{n,p}$ (see [59]), which was ultimately proved using the hypergraph container method in [4] and [71]. An embedding lemma is weaker than a counting lemma, only providing the existence of at least one appearance of some subgraph H, whereas a counting lemma provides roughly the expected number of copies based on edge densities between parts of a vertex partition. Stronger "probabilistic counting lemmas" for $\boldsymbol{G}^{(2)}_{n,p}$ were obtained in [28] and [43], motivated in particular by Turán-type theorems (as discussed in Section 2.1). We note an interesting contrast: whereas these works make use of the deterministic Kohayakawa–Rödl sparse regularity lemma (with a *no-dense-spots* condition) and establish a counting lemma that holds for dense subgraphs of $\boldsymbol{G}^{(2)}_{n,p}$ with high probability, here we combine a deterministic counting lemma with a decomposition lemma holding (with acceptable complexity for our application) with high probability.

2.3. *The* B^**-norms*

There are several natural generalizations of the cut norm to r-tensors. For instance, one can take cut tensors of the form $\mathbf{1}_{I_1} \otimes \cdots \otimes \mathbf{1}_{I_r}$ for $I_1, \dots, I_r \subset [n]$, as was done for the sparse Frieze–Kannan decomposition proved in [23]. However, it turns out that the resulting norm is only useful for controlling homomorphism densities for linear hypergraphs H (see [64]).

Instead, we will consider the wider class of Boolean tensors formed by entry-wise products of tensors varying on strict subsets of the r coordinates. We have the following.

Definition 2.5 (Weighted base)
Given a set e of size r, a *base (over e)* is a collection F of proper subsets of e with $\emptyset \in \mathsf{F}$ and such that any two nonempty elements $\mathsf{f}_1, \mathsf{f}_2 \in \mathsf{F}$ are incomparable, that is, $\mathsf{f}_1 \not\subset \mathsf{f}_2$. A *weighted base* over e is a tuple $\mathsf{B} = (r, e, \iota, \mathsf{F}, d_\star, \{d_\mathsf{f}\}_{\mathsf{f}\in\mathsf{F}})$, with

- $\iota : [r] \to e$ a bijective mapping;

- F a base over e;
- nonnegative integer weights $d_\star$ and d_f satisfying $d_\mathsf{f} \le d_\star$ for each $\mathsf{f} \in \mathsf{F}$, and $d_\emptyset := 0$.

A *base system* over an r-graph H is a collection $\mathbb{B} = \{\mathsf{B}(e)\}_{e \in \mathsf{E}(H)}$ with each $\mathsf{B}(e)$ a weighted base over e.

In our applications to H-counts, the choice of integer weights $d_\star$, d_f will generally be determined by the degrees of an edge and its subsets.

For a weighted base B we define an associated set of *test tensors* $\mathcal{T}_\mathsf{B} \subset \mathcal{Z}_{n,r}$ consisting of all nonzero Boolean tensors $T : [n]^r \to \{0,1\}$ of the form

$$T(i_1, \dots, i_r) = \prod_{\mathsf{f} \in \mathsf{F}} \tau_\mathsf{f} \circ \pi_\mathsf{f}(i_1, \dots, i_r), \quad i_1, \dots, i_r \in [n], \tag{2.10}$$

for general Boolean functions $\tau_\mathsf{f} : [n]^\mathsf{f} \to \{0,1\}$, where we denote by $\pi_\mathsf{f} : [n]^r \to [n]^\mathsf{f}$ the natural projections $(i_1, \dots, i_r) \mapsto (i_v)_{v \in \iota^{-1}(\mathsf{f})}$. We always take $\tau_\emptyset$ to be the constant tensor $\tau_\emptyset(i_1, \dots, i_r) \equiv 1$. We quantify the size of the factors of a test tensor as in (2.10) with the rescaled $\ell_1([n]^\mathsf{f})$-norm:

$$\|T\|_\mathsf{f} := n^{r-|\mathsf{f}|} p^{d_\star - d_\mathsf{f}} \|\tau_\mathsf{f}\|_1 \tag{2.11}$$

(in particular $\|T\|_\emptyset = n^r p^{d_\star}$), and define

$$\|T\|_\mathsf{B} := \max\Big\{\|T\|_1, \max_{\mathsf{f} \in \mathsf{F}} \|T\|_\mathsf{f}\Big\}. \tag{2.12}$$

(This can be extended to a norm on $\mathcal{Z}_{n,r}$, but we only apply it to test tensors.) Note that the inclusion of $\emptyset \in \mathsf{F}$ means that we always have $\|T\|_\mathsf{B} \ge n^r p^{d_\star}$. We define a seminorm $\|\cdot\|_\mathsf{B}^*$ on $\mathcal{Z}_{n,r}$ via duality:

$$\|Z\|_\mathsf{B}^* := \max_{T \in \mathcal{T}_\mathsf{B}} \frac{|\langle Z, T\rangle|}{\|T\|_\mathsf{B}}. \tag{2.13}$$

When F covers $[r]$ this defines a genuine norm on $\mathcal{Z}_{n,r}$, but we do not enforce this in general. We note that these seminorms additionally depend on all components of the weighted base B (not just the base F) as well as p, but we suppress this dependence from the notation. If $\mathbb{B} = \{\mathsf{B}(e)\}_{e \in \mathsf{E}}$ is a finite collection of weighted bases $\mathsf{B}(e)$ over the respective r-sets e (such as a base system over an r-graph H, with $\mathsf{E} = \mathsf{E}(H)$), then we denote the seminorm

$$\|Z\|_\mathbb{B}^* := \max_{e \in \mathsf{E}} \|Z\|_{\mathsf{B}(e)}^*. \tag{2.14}$$

Example 2.6

For the case $r = 2$, we recover the matrix norm (2.3) by taking the maximal base

$\mathsf{F} = \{\emptyset, \{1\}, \{2\}\}$ over $\{1, 2\}$, $d_\star = 2\Delta - 2$ and $d_{\{1\}} = d_{\{2\}} = \Delta - 1$, and this further specializes to the normalized cut norm $\frac{1}{n^2}\|\cdot\|_\square$ upon taking $p = 1$.

For general $r \geq 2$, it is useful to consider the maximal base $\mathsf{F} = \binom{[r]}{r-1} \cup \{\emptyset\}$ with appropriate degree parameters. For instance, we have the following.

Example 2.7
In the $K_4^{(3)}$-counting lemma presented in Theorem 1.5, the bound (1.11) is equivalent up to constant factors to the bound $\|A_{G_1} - A_{G_2}\|_{\mathsf{B}}^* \leq \varepsilon p$, where for the weighted base B we take $e = \{1, 2, 3\}$, ι the identity, $\mathsf{F} = \{\emptyset, \{1, 2\}, \{2, 3\}, \{1, 3\}\}$, $d_\star = 3$, and $d_{\{1,2\}} = d_{\{2,3\}} = d_{\{1,3\}} = 1$. Indeed, for this choice of base a test tensor takes the form $\mathbf{1}_{I^{(2,3)} \cap J^{(1,3)} \cap K^{(1,2)}}$ for sets $I, J, K \subset [n]^2$. (The equivalence is up to a constant because we take a sum on the right-hand side in (1.11) rather than a maximum as in (2.12).)

For general r and F and with $p = 1$, we have $\|T\|_{\mathsf{B}} = n^r$ and we recover the family of generalized cut norms considered by Conlon and Lee in [30] (various special cases of which had been considered earlier, such as the case $\mathsf{F} = \binom{[r]}{r-1} \cup \{\emptyset\}$ by Gowers in [44], [45]). It was shown in [30] for the unweighted setting ($p = 1$) that $\|\cdot\|_{\mathsf{B}}^*$ is polynomially equivalent to certain generalized Gowers norms, though the polynomial loss appears to make the latter ineffective in the sparse setting (this is related to how various definitions of quasirandomness for dense graphs cease to be equivalent for sparse graphs).

For our application to H-homomorphism counts (through the counting lemma, Theorem 2.15 below), for each edge $e \in \mathsf{E}(H)$ we will select a weighted base over e satisfying the following property.

Definition 2.8 (Dominating base)
For an r-graph H and $e \in \mathsf{E}(H)$, a base F over e is *H-dominating* if every edge overlap $e \cap e'$ with $e' \neq e$ is contained in some $\mathsf{f} \in \mathsf{F}$. A weighted base is H-dominating if its base is H-dominating. A base system $\mathbb{B} = \{\mathsf{B}(e)\}_{e \in \mathsf{E}(H)}$ over H is H-dominating if each of the weighted bases $\mathsf{B}(e)$ is H-dominating.

Observe that if the weighted bases $\mathsf{B} = (r, e, \iota, \mathsf{F}, d_*, \{d_\mathsf{f}\}_{\mathsf{f} \in \mathsf{F}})$ and $\mathsf{B}' = (r, e, \iota, \mathsf{F}', d_*, \{d'_\mathsf{f}\}_{\mathsf{f}' \in \mathsf{F}'})$ over bases F and F', respectively, are such that $d_* = d'_*$, any $\mathsf{f}' \in \mathsf{F}'$ is contained in F, and $d_{\mathsf{f}'} = d_\mathsf{f}$, then $\|T\|_{\mathsf{B}} \geq \|T\|_{\mathsf{B}'}$. Thus $\|Z\|_{\mathsf{B}}^* \leq \|Z\|_{\mathsf{B}'}^*$.

With a fixed choice of dominating base $\mathsf{F}(e)$ for each $e \in \mathsf{E}(H)$, and some arbitrary choice of bijections $\iota_e : [r] \to e$, we take the H-dominating base system $\mathbb{B} = \{\mathsf{B}(e)\}_{e \in \mathsf{E}(H)}$ with

$$\mathsf{B}(e) = \left(r, e, \iota_e, \mathsf{F}(e), d^H(e), \left\{d_{\mathsf{f}}^H(e)\right\}_{\mathsf{f}\in\mathsf{F}(e)}\right) \tag{2.15}$$

and with degree parameters as defined in (1.18) and (1.19). The choice of integer weights is motivated by the proof of the counting lemma, where the prefactors in (2.11) combined with the hypothesis of a crude upper bound on counts of subgraphs of H (as in (1.14)) will allow us to close an induction on the number of edges in H. (We only depart from the choice of weights (2.15) in the *proof* of Theorem 2.15, via the generalization Theorem 4.1, where we take H-dominating bases with weights $d^{H+}(e)$, $d_{\mathsf{f}}^{H+}(e)$ taken according to a subgraph H_+ of H, but in the statement of the theorem, and hence in all of its applications, we take weights as in (2.15).)

Example 2.9
Continuing the example of Theorem 1.5, the base of Example 2.7 can be made into a base system $\mathbb{B}$ over the edges of $K_4^{(3)}$ by taking the weighted base B for each edge (identifying each with $\{1,2,3\}$), and one verifies that the weights $d_\star$, d_{f} given in Example 2.7 are chosen as in (2.15).

While the maximal base $\mathsf{F} = \{\emptyset\} \cup \binom{e}{r-1}$ (as in Examples 2.6 and 2.7) is always dominating, for certain H one can use bases with a smaller number of smaller sets, which leads to better quantitative estimates. For instance, we have the following.

Example 2.10
When H is a *sunflower*, where pairwise overlaps of all edges are equal to a common *kernel* $V_0 \subseteq \mathsf{V}(H)$ (thus $\Delta(H) = \mathsf{e}(H)$), then for any $e \in \mathsf{E}(H)$ one can take the dominating base $\mathsf{F} = \{\emptyset, V_0\}$, with weights as in (2.15) being $d_\star = d_{V_0} = \Delta(H) - 1$. With this choice Theorem 2.13 gives a more efficient approximation of the adjacency tensor $\boldsymbol{A}$ by test tensors, leading to a less restrictive decay condition on p (see (2.16)).

Example 2.11
For the case of 2-graphs, for an edge $e = \{u, v\}$ one can always take $\mathsf{F}(e) = \{\emptyset, \{u\}, \{v\}\}$ as in Example 2.6. If u (resp., v) has degree 1, then one can take the smaller dominating base $\mathsf{F}(e) = \{\emptyset, \{v\}\}$ (resp., $\{\emptyset, \{u\}\}$). In this case, the adjacency matrices for two graphs are close in the $\mathsf{B}(e)^*$-(semi)norm when they have approximately the same degree sequence. Such an approximation is indeed sufficient for approximating $\hom(H, G)$ when H is a star (for which we can take such a base for every edge) as these are moments of the degree distribution.

Example 2.12
When H is a linear hypergraph, for which $|e \cap e'| \in \{0, 1\}$ for all distinct $e, e' \in \mathsf{E}(H)$,

then for each edge e the base of singletons $\mathsf{F}(e) = \{\emptyset\} \cup \{\{v\} : v \in e\}$ is dominating, and the class of test tensors thus reduces to the class of cuts $\mathbf{1}_{I_1} \otimes \cdots \otimes \mathbf{1}_{I_r}$.

We return to these and other examples in Section 3.2, where we state results for tails of homomorphism counts in $\boldsymbol{G}_{n,p}$.

2.4. B^* *decomposition and counting lemmas*

The following is our general decomposition lemma, showing that, under the Erdős–Rényi measure, most symmetric Boolean tensors $A \in \mathscr{A}_{n,r}$ can be decomposed into a structured piece A_{str} that is a combination of a small number of test tensors of controlled size under the B-norm, and a pseudorandom piece A_{rand} that is small in the dual B^*-norm. The theorem allows for a trade-off between the measure of the set of tensors to be excluded and the complexity of the resulting decomposition (in particular one may choose to make no exclusion). Recall our notation $J_{n,r}$ for the symmetric Boolean r-tensor with $J_{n,r}(i_1, \dots, i_r) = 1$ if and only if all of the arguments $i_1, \dots, i_r$ are distinct (so that for the Erdős–Rényi tensor we have $\mathbb{E}\boldsymbol{A} = pJ_{n,r}$).

THEOREM 2.13 (Decomposition lemma)
The following holds for $C_0, c_0 > 0$ depending only on r. Fix a weighted base $\mathsf{B} = (r, e, \iota, \mathsf{F}, d_\star, \{d_{\mathsf{f}}\}_{\mathsf{f}\in\mathsf{F}})$, and let $\kappa, \varepsilon > 0$. Assuming n and $p \in (n^{-2}, 1)$ are such that

$$w_{n,p}(\mathsf{B}) := \min_{\mathsf{f}\in\mathsf{F}} \{n^{r-|\mathsf{f}|} p^{d_\star - d_{\mathsf{f}} + 2}\} \geq \frac{C_0 \log n}{\varepsilon^2 (1 \vee \log(1/p))}, \tag{2.16}$$

then there exists a (possibly empty) exceptional set $\mathcal{E}_\star(\kappa, \varepsilon) \subseteq \mathscr{A}_{n,r}$ with $\mathbb{P}(\boldsymbol{A} \in \mathcal{E}_\star(\kappa, \varepsilon)) \leq p^{c_0 \kappa n^r}$ such that for each $A \in \mathscr{A}_{n,r} \setminus \mathcal{E}_\star(\kappa, \varepsilon)$ there is a decomposition

$$A = A_{\mathrm{str}} + A_{\mathrm{rand}}, \tag{2.17}$$

where

$$A_{\mathrm{str}} = pJ_{n,r} + \sum_{i=1}^{k} \alpha_i T_i \qquad \text{and} \qquad \|A_{\mathrm{rand}}\|_{\mathsf{B}}^* \leq \varepsilon p \tag{2.18}$$

for real numbers $\alpha_1, \dots, \alpha_k$ and test tensors $T_1, \dots, T_k \in \mathcal{T}_{\mathsf{B}}$ satisfying

$$\sum_{i=1}^{k} \|T_i\|_{\mathsf{B}} \leq \kappa n^r / (\varepsilon p)^2. \tag{2.19}$$

Furthermore, for each $1 \leq j \leq k$, the tensor T_j is separated from the span of $\{T_1, \dots, T_{j-1}\}$ by Euclidean distance at least $\varepsilon p^{1+d_\star} n^{r/2}$.

Remark 2.14
The final statement on Euclidean distances will be useful for bounding covering numbers of $\mathcal{A}_{n,r} \setminus \mathcal{E}_\star(\kappa, \varepsilon)$ under the B^*-norm.

The following is our general sparse counting lemma, which we state for a multilinear generalization of homomorphism counts. For a collection $\underline{S}$ of symmetric tensors $\{S^e\}_{e \in \mathsf{E}(H)}$ in $\mathcal{S}_{n,r}$, we define

$$t_p(H, \underline{S}) = \frac{\hom(H, \underline{S})}{n^{\mathsf{v}(H)} p^{\mathsf{e}(H)}} := \frac{1}{n^{\mathsf{v}(H)} p^{\mathsf{e}(H)}} \sum_{\phi : \mathsf{V}(H) \to [n]} \prod_{e \in \mathsf{E}(H)} S^e\big(\phi(e)\big).$$

For $\underline{S}$ with $S^e \equiv S_0$ for some $S_0 \in \mathcal{S}_{n,r}$ the above expression reduces to the previous definition $t_p(H, \underline{S}) = t_p(H, S_0)$ from (1.21). This multilinear generalization of homomorphism counts allows us to capture other functionals of interest such as induced subgraph counts. It also naturally appears in the proof of the counting lemma, which interpolates between two (weighted) hypergraphs that are close under the $\|\cdot\|_{\mathbb{B}}^*$-seminorm.

Recall the notion of a dominating base system from Definition 2.8.

THEOREM 2.15 (Counting lemma)
Let $p \in (0,1)$, and let H be an r-graph. Let $\mathbb{B}$ be an H-dominating base system as in (2.15), and let $\|\cdot\|_{\mathbb{B}}^*$ be the associated seminorm on $\mathcal{Z}_{n,r}$ as defined in (2.14). Let $\mathcal{C} \subset \mathcal{A}_{n,r}$ be a set of diameter at most εp under $\|\cdot\|_{\mathbb{B}}^*$ for some $\varepsilon \in (0,1]$, and assume further that there exist $Q_0 \in \mathrm{hull}(\mathcal{C})$ and $L \geq 1$ such that

$$t_p(H', Q_0) \leq L \tag{2.20}$$

for all proper subgraphs $H' \subset H$. Then for any $\underline{P} = \{P^e\}_{e \in \mathsf{E}(H)}$, $\underline{Q} = \{Q^e\}_{e \in \mathsf{E}(H)}$ with each $P^e, Q^e \in \mathrm{hull}(\mathcal{C})$, we have

$$t_p(H, \underline{P}) - t_p(H, \underline{Q}) \lesssim_H L\varepsilon.$$

The above is a consequence of Theorem 4.1 giving a counting lemma for the broader class of *signed-homomorphism* functionals interpolating between homomorphism counts and *induced* homomorphism counts.

3. Quantitative LDPs

3.1. Quantitative B^*-norm LDPs

As a consequence of Theorem 2.13, we obtain quantitative LDPs for the measure space $(\mathcal{A}_{n,r}, \mu_p)$ at large fixed n. For comparison, the classical LDP for a sequence of measures ν_n on a topological space $\mathcal{Q}$ states that for any $\mathcal{E} \subset \mathcal{Q}$,

$$-\inf_{x\in\mathcal{D}} I(x) \le \liminf_{n\to\infty} \frac{1}{R_n}\log \nu_n(\mathcal{E}) \le \limsup_{n\to\infty} \frac{1}{R_n}\log \nu_n(\mathcal{E}) \le -\inf_{x\in\mathcal{F}} I(x) \tag{3.1}$$

for any open $\mathcal{D}\subseteq\mathcal{E}$ and closed $\mathcal{F}\supseteq\mathcal{E}$, where R_n is the speed and $I(\cdot)$ is the LDP rate function.

For our quantitative result, the rate function is the relative entropy $\mathrm{I}_p(Q) = \mathrm{D}(\mu_Q\|\mu_p)$ (see (1.17)). Now we specify our notions of outer and inner approximations of a set $\mathcal{E}\subset\mathcal{A}_{n,r}$. Let E be an arbitrary finite collection of r-sets, and let $\mathbb{B} = \{\mathsf{B}(e)\}_{e\in\mathsf{E}}$ be a collection of weighted bases $\mathsf{B}(e)$ over e. Recalling the associated seminorm defined in (2.14), for $Q\in\mathcal{Q}_{n,r}$, $\delta>0$, we denote the δ-neighborhood of Q in $\mathcal{A}_{n,r}$ by

$$\mathcal{U}_{\mathbb{B}}(Q,\delta) := \big\{A\in\mathcal{A}_{n,r} : \|Q-A\|_{\mathbb{B}}^{*}\le\delta\big\}. \tag{3.2}$$

For $\mathcal{E}\subseteq\mathcal{A}_{n,r}$, we denote the outer approximation

$$(\mathcal{E})_{\mathbb{B},\delta} := \bigcup_{A\in\mathcal{E}} \operatorname{hull}\big(\mathcal{U}_{\mathbb{B}}(A,\delta)\big) \tag{3.3}$$

and the inner approximation

$$(\mathcal{E})^{\circ}_{\mathbb{B},\delta} := \big\{Q\in\mathcal{Q}_{n,r} : \mathcal{U}_{\mathbb{B}}(Q,\delta)\subseteq\mathcal{E}\big\}. \tag{3.4}$$

Note that $(\mathcal{E})_{\mathbb{B},\delta}$ and $(\mathcal{E})^{\circ}_{\mathbb{B},\delta}$ are subsets of the solid cube $\mathcal{Q}_{n,r}$. We also emphasize that the convex hulls in (3.3) are different from (and generally proper subsets of) the balls $\{Q\in\mathcal{Q}_{n,r} : \|Q-A\|_{\mathbb{B}}^{*}\le\delta\}$.

THEOREM 3.1 (Quantitative LDP)
Let $\mathbb{B}$ be a collection of weighted bases as above.

(a) *(LDP upper bound). With C_0, c_0 as in Theorem 2.13, let $p\in(n^{-2},1)$, $\kappa,\varepsilon>0$, and assume that $w_{n,p}(\mathbb{B}) := \min_{e\in\mathsf{E}} w_{n,p}(\mathsf{B}(e))$ satisfies the lower bound in (2.16). Then for any $\mathcal{E}\subseteq\mathcal{A}_{n,r}$,*

$$\log\mathbb{P}(\boldsymbol{A}\in\mathcal{E}) \le -\min\Big(R_\star, \inf\big\{\mathrm{I}_p(Q) : Q\in(\mathcal{E})_{\mathbb{B},\varepsilon p}\big\} - R_{\mathrm{ME}}\Big) \tag{3.5}$$

for some cutoff rate

$$R_\star = c_0\kappa n^r\log(1/p) - O_{\mathbb{B}}(1) \tag{3.6}$$

and metric entropy rate

$$R_{\mathrm{ME}} \lesssim_{\mathbb{B}} \frac{\kappa n^r\log n}{\varepsilon^2 w_{n,p}(\mathbb{B})}. \tag{3.7}$$

(b) *(LDP lower bound). If* $w_{n,p}(\mathbb{B}) \geq C_0' \varepsilon^{-2} \log n$ *for a sufficiently large constant* $C_0'(r) > 0$*, then for any* $\mathcal{E} \subseteq \mathcal{A}_{n,r}$*,*

$$\log \mathbb{P}(\boldsymbol{A} \in \mathcal{E}) \geq -\inf\{\mathrm{I}_p(Q) : Q \in (\mathcal{E})^{\circ}_{\mathbb{B},\varepsilon p}\} - O\big(1 + n^{r/2}\big|\log\big(p \wedge (1-p)\big)\big|\big), \tag{3.8}$$

where the implied constant is absolute.

In applications, we take κ such that $R_\star$ exceeds the rate of the rare event of interest—for the upper tail of $t_p(H, \boldsymbol{G})$ this means taking $\kappa = Kp^{\Delta}$ for a sufficiently large constant K. From our assumption on $w_{n,p}(\mathbb{B})$ we have $R_{\mathrm{ME}} = O(R_\star)$, but we must further ensure that $w_{n,p}(\mathbb{B}) \gg 1$ to have R_{ME} be negligible compared to the main term. This amounts to a lower-bound constraint on p, and there is generally flexibility (within the requirements of a counting lemma) to choose the weighted bases to lighten this constraint.

The upper bound of Theorem 3.1 follows from Theorem 2.13 by a straightforward covering argument, combined with the nonasymptotic bound (6.1). The term R_{ME} is the sum of log-covering numbers of $\mathcal{A}_{n,r} \setminus \mathcal{E}_\star(Kp^{\Delta}, \varepsilon)$ by εp-balls in the $\mathsf{B}(e)^*$-norms. (The refinement to convex hulls of their intersections is important when combining Theorem 3.1 with Theorem 2.15.) To obtain bounds for upper tails of functionals $f : \mathcal{Q}_{n,r} \to \mathbb{R}$ we apply Theorem 3.1 to $\mathcal{E} = \{f \geq t\}$, and then use Theorem 2.15 to show that $\{f \geq t + \eta\} \subset (\mathcal{E})^{\circ}_{\mathbb{B},\varepsilon p} \subset (\mathcal{E})_{\mathbb{B},\varepsilon p} \subset \{f \geq t - \eta\}$ for some $\eta = o_{\varepsilon \to 0}(1)$.

3.2. Upper and lower tails for hypergraph counts

Our main application of Theorem 3.1, in combination with the counting lemma (Theorem 2.15), is Theorem 1.1 on joint upper and lower tails for homomorphism counts. In this subsection, we state some corollaries of Theorem 1.1 for specific classes of hypergraphs and define the parameter $\Delta'(H)$. Further applications of Theorems 3.1 and 2.15 are given in Section 10.

For the case $m = 1$ and $p = o(1)$, the optimization problem $\Phi_{n,p}(H, \delta)$ was recently analyzed in [61] for certain H—specifically, complete hypergraphs and the 3-graph depicted in Figure 1—where they deduced upper-tail asymptotics for $p \gg n^{-1/(6\mathsf{e}(H))} \log n$ via the general framework from [38], which required bounding the Gaussian width of gradients for the homomorphism counting functionals. Combining Theorem 1.1 with [61, Theorem 2.3], we obtain the following.

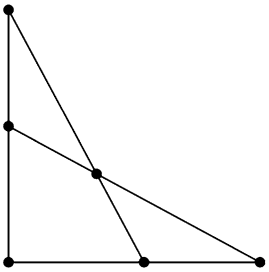

Figure 1. The 3-graph considered in [61] (see Corollary 3.2) has six vertices (dots) with four edges denoted by straight lines. (Reproduced with permission from [61].)

COROLLARY 3.2
For fixed r, k, and $K_k^{(r)}$ the r-uniform clique on k vertices, we have

$$\begin{aligned}&\log \mathbb{P}\big(t_p(K_k^{(r)}, \boldsymbol{G}) \geq 1+\delta\big)\\ &\quad = -\big(1+o(1)\big)\min\Big\{\frac{\delta^{r/k}}{r!}, \frac{\delta}{(r-1)!k}\Big\} n^r p^{\binom{k-1}{r-1}} \log(1/p) \qquad (3.9)\end{aligned}$$

if $n^{-c(r,k)} \ll p \ll 1$ with $c(r,k) = 1/(\binom{k-1}{r-1}+1)$. Furthermore, the lower bound holds for the wider range $n^{-1/\binom{k-1}{r-1}} \ll p \ll 1$.

Moreover, with H the 3-graph depicted in Figure 1, for $n^{-1/2} \ll p \ll 1$, we have

$$\begin{aligned}&\log \mathbb{P}\big(t_p(H, \boldsymbol{G}) \geq 1+\delta\big)\\ &\quad = -\Big(\frac{1}{6}+o(1)\Big)\min\{\sqrt{9+3\delta}-3, \sqrt{\delta}\} n^3 p^2 \log(1/p). \qquad (3.10)\end{aligned}$$

The ranges of p for the upper bounds follow from our computation of the parameters $\Delta'(K_k^{(r)})$ and $\Delta'(H)$ in Examples 3.6 and 3.8 below. Analogously to the case $r = 2$, the asymptotic (3.9) for cliques matches the probability of appearance of higher-rank analogues of the "clique" and "hub" structures that were first described in [65]. One may view H from Figure 1 as the 3-graph obtained by transposing the incidence matrix of the complete 2-graph $G = K_4$. The interest in this particular hypergraph is that the mechanism for large deviations of $\hom(H, \boldsymbol{G})$ is more intricate than the simple appearance of a clique or hub structure as is the case for $H = K_k^{(r)}$ (see [61] for further discussion).

We next highlight consequences of Theorem 1.1 combined with a result from [9] providing an asymptotic for $\Phi_{n,p}(H, \delta)$ for the for the case $r = 2$, $m = 1$, and $p = o(1)$.

COROLLARY 3.3 (The case of 2-graphs)
Let H be a fixed 2-graph of maximal degree $\Delta \geq 2$.

(a) *For any fixed $\delta > 0$, assuming $n^{-1/(\Delta+1)} \ll p \leq (1+\delta)^{-1/\mathsf{e}(H)}$,*

$$\log \mathbb{P}\big(t_p(H, \boldsymbol{G}) \geq 1+\delta\big) = -\big(1+o(1)\big)\Phi_{n,p}\big(H, \delta + o(1)\big) \tag{3.11}$$

and for fixed $\delta \in (0,1)$ and $n^{-1/(\Delta+1)} \log n \ll p < 1$,

$$\log \mathbb{P}\big(t_p(H, \boldsymbol{G}) \leq 1-\delta\big) = -\big(1+o(1)\big)\Psi_{n,p}\big(H, \delta + o(1)\big). \tag{3.12}$$

(b) *Furthermore, (3.11) (resp., (3.12)) holds in the range $n^{-1/\Delta} \ll p \leq (1+\delta)^{-1/\mathsf{e}(H)}$ (resp., $p \gg n^{-1/\Delta}\log n$) whenever each vertex of H of degree Δ is contained in an isolated star.*

(c) *Further specializing to the case that $H = K_{1,\Delta}$ (the Δ-armed star) and $n^{-1/\Delta} \ll p \ll 1$,*

$$\log \mathbb{P}\big(t_p(K_{1,\Delta}, \boldsymbol{G}) \geq 1+\delta\big) = -\big(1+o(1)\big)\delta n^2 p^{\Delta} \log(1/p). \tag{3.13}$$

Remark 3.4
For general r, the asymptotic (3.11) holds in the range $n^{-1/\Delta(H)} \ll p \leq (1+\delta)^{-1/\mathsf{e}(H)}$, for instance, when H is a sunflower with at least $2/(r - |V_0|)$ petals, where $V_0 \subset \mathsf{V}(H)$ is the kernel of H (see Example 3.7).

Remark 3.5
The assumption on p in part (c) is sharp, as the upper-tail rate is known to be of size $\Theta_\delta(n^{1+1/\Delta} p \log(1/p))$ for $n^{-(1+1/\Delta)}(\log n)^{1/(\Delta-1)} \lesssim p \lesssim n^{-1/\Delta}$ (see [74]).

Proof
Part (a) is immediate from Theorem 1.1. For (b), we only need to verify that $\Delta'(H) \leq \Delta(H)$ for such 2-graphs, which we do in Example 3.9 below. Part (c) follows from Theorem 1.1 and [9, Theorem 1.5], where we note that the independence polynomial of the graph H^* defined in that work is $P_{H^*}(\theta) = 1+\theta$ whenever H is a star. □

We now define the parameter $\Delta'(H)$ appearing in Theorem 1.1. Roughly speaking, it measures how efficiently one can cover the edge overlaps of H, with a small value for r-graphs with overlaps concentrated on a small number of small vertex sets (such as sunflowers or sparse linear hypergraphs) and a large value for cliques. A key point is that it depends only on the neighborhood structure of single edges, similarly to how $\Delta(H)$ depends only on the neighborhood of single vertices. Thus it is a *local* hypergraph parameter that is independent of the size of H (as quantified by $\mathsf{v}(H)$ or $\mathsf{e}(H)$).

Recall the notion of a dominating weighted base from Definition 2.8. Given a dominating base F over an edge $e \in \mathsf{E}(H)$, recalling the edge degree parameters from (2.15), we set

$$d'_{\mathsf{F}}(e) = \max\bigl\{d'_{\mathsf{f}}(e) : \mathsf{f} \in \mathsf{F}\bigr\},$$
$$d'_{\mathsf{f}}(e) = \frac{d^H(e) - d^H_{\mathsf{f}}(e) + 2}{|e \setminus \mathsf{f}|} = \begin{cases} \frac{d^H(e\setminus \mathsf{f})+1}{|e\setminus \mathsf{f}|} & \mathsf{f} \neq \emptyset, \\ \frac{d^H(e)+2}{r} & \mathsf{f} = \emptyset. \end{cases} \tag{3.14}$$

(Recall that $d^H(e)$ does not count e itself.) Note that $d'_{\mathsf{f}}(e)$ is a normalized count of the edges overlapping $e \setminus \mathsf{f}$, that is, those that are *not* dominated by f. (If the 2 were replaced by 1 in the first expression for $d'_{\mathsf{f}}(e)$, then it would be the average degree of vertices in $e \setminus \mathsf{f}$.) We define

$$d'(e) = \min_{\mathsf{F}} d'_{\mathsf{F}}(e), \qquad \Delta'(H) = \max_{e \in \mathsf{E}(H)} d'(e), \tag{3.15}$$

where the minimum is taken over all dominating bases F for e. That is, $\Delta'(H)$ is the smallest number Δ' such that every edge $e \in \mathsf{E}(H)$ has a dominating base $\mathsf{F}(e)$ such that

$$\bigl|\bigl\{e' \in \mathsf{E}(H) : e' \cap (e \setminus \mathsf{f}) \neq \emptyset\bigr\}\bigr| \leq \Delta' |e \setminus \mathsf{f}| - 1 \tag{3.16}$$

for every $\mathsf{f} \in \mathsf{F}(e)$ (note that the left-hand side counts the edge e). From this definition, recalling the notation of (2.16) and Theorem 3.1, we have that for any r-graph H, whenever $np^{\Delta'(H)} \geq 1$, there exists an H-dominating base system $\mathbb{B} = \{\mathsf{B}(e)\}_{e \in \mathsf{E}(H)}$ with weights as in (2.15) such that

$$w_{n,p}(\mathbb{B}) = \min_{e \in \mathsf{E}(H)} w_{n,p}\bigl(\mathsf{B}(e)\bigr) = \min_{e \in \mathsf{E}(H)} \min_{\mathsf{f} \in \mathsf{F}(e)} (np^{d'_{\mathsf{f}}(e)})^{r-|\mathsf{f}|} \geq np^{\Delta'(H)}. \tag{3.17}$$

We record some general bounds on $\Delta'(H)$ by considering specific dominating bases. We drop the superscript H from all notation for the remainder of this section. We write

$$\Delta_\star(H) = \max_{e \in \mathsf{E}(H)} d(e)$$

for the maximal edge degree (which does not count the edge e itself). For $1 \leq s \leq r$, let

$$\Delta_s(H) = \max_{e \in \mathsf{E}(H)} \max_{U \in \binom{e}{s}} \bigl|\{e' : U \cap e' \neq \emptyset\}\bigr|$$

denote the largest number of hyperedges intersecting a size-s subset of some edge of H; in particular, $\Delta_1(H) = \Delta(H)$, $\Delta_r(H) = \Delta_\star(H) + 1$, and $\Delta_s(H) \leq s\Delta(H)$ for every $1 \leq s \leq r-1$. Since $d'(e) \geq (d(e)+2)/r$, it follows that for any r-graph H,

$$\Delta'(H) \geq \frac{\Delta_\star(H)+2}{r} \geq \frac{\Delta(H)+1}{r}, \tag{3.18}$$

whereas taking $\mathsf{F}(e) = \binom{e}{r-1} \cup \{\emptyset\}$ (which is always a dominating base) shows that

$$\Delta'(H) \leq \Delta(H)+1. \tag{3.19}$$

Indeed, for each $\mathsf{f} = e \setminus \{v\}$ we have $d(e \setminus \mathsf{f}) = d(v) \leq \Delta(H)$. If every pair of edges overlaps in at most s_0 vertices, then taking the bases $\mathsf{F}(e) = \{\emptyset\} \cup \binom{e}{s_0}$ we obtain the sharper bound

$$\Delta'(H) \leq \max_{s \in \{r, r-s_0\}} \frac{\Delta_s(H)+1}{s}. \tag{3.20}$$

Example 3.6 (Cliques)
When H is the r-uniform clique on k vertices, we have $\Delta(H) = \binom{k-1}{r-1}$ and $\Delta'(H) = \binom{k-1}{r-1}+1$, so that equality holds in (3.19). Indeed, for each hyperedge e we are forced to take $\mathsf{F}(e) = \binom{e}{r-1} \cup \{\emptyset\}$ to satisfy the domination condition.

Example 3.7 (Sunflowers and stars)
When H is a sunflower, with pairwise intersections of all Δ edges ("petals") equal to a common "kernel" $V_0 \subset \mathsf{V}(H)$, for every edge the optimal base is $\mathsf{F}(e) = \{\emptyset, V_0\}$, for which we have $d'_\emptyset(e) = \frac{\Delta+1}{r}$ and $d'_{V_0}(e) = \frac{2}{r-|V_0|}$, and so $\Delta'(H) = \max\{\frac{\Delta+1}{r}, \frac{2}{r-|V_0|}\}$. Thus, sunflowers attain the minimum in (3.18) as long as the kernel is of size $|V_0| \leq r\frac{\Delta-1}{\Delta+1}$. For the r-uniform Δ-armed star, with $\Delta \geq 2$ (with $|V_0| = 1$) we have $\Delta'(H) = \frac{\Delta+1}{r}$ when $r \geq 3$ and $\Delta'(H) = \max\{\frac{\Delta+1}{2}, 2\} \leq \Delta$ when $r = 2$.

Example 3.8 (Linear hypergraphs)
When all pairs of edges share at most one vertex, then (3.20) holds with $s_0 = 1$. For instance, for linear cycles (or disjoint unions thereof), $\Delta'(H) = \max\{\frac{4}{r}, \frac{3}{r-1}\}$ which attains the lower bound (3.18) of $4/r$ for all $r \geq 4$. For 2-graphs of degree 2 we have $\Delta'(H) \leq 3$, and one checks that in fact $\Delta'(H) = 3$. For the linear 3-graph of Corollary 3.2, taking $\mathsf{F}(e) = \binom{e}{1} \cup \{\emptyset\}$ shows that $\Delta'(H) = \frac{1}{2}(\Delta_2(H)+1) = 2$. For the Fano plane, one checks that $\Delta'(H) = \Delta(H) = 3$.

Example 3.9 (2-graphs)
As was noted in (3.19), we always have $\Delta'(H) \leq \Delta(H)+1$. Here we verify that $\Delta'(H) \leq \Delta(H)$ for any 2-graph H as in Corollary 3.3(b). Indeed, for any edge $e = \{u, v\}$ with an end v of maximal degree we take the base $\mathsf{F}(e) = \{\emptyset, \{v\}\}$, giving $d'(e) = \max\{\frac{1}{2}(\Delta(H)+1), 2\} \leq \Delta(H)$. For any other edge, the maximal base

$\mathsf{F}(e) = \{\emptyset, \{u\}, \{v\}\}$ verifies that $d'(e) \le \Delta(H)$, and thus $\Delta'(H) = \max_e d'(e) \le \Delta(H)$.

Remark 3.10
As in Remark 2.3, we believe that the right-hand side of (2.19) can be improved to $\kappa n^r \epsilon^{-2} p^{-1}$, which would allow us to replace the left-hand side of (2.16) with $\min_{\mathsf{f}\in\mathsf{F}}\{n^{r-|\mathsf{f}|} p^{d_*-d_f+1}\}$. This would in turn imply that the conclusion of Theorem 1.1 holds as long as $p \gg n^{-1/\tilde{\Delta}(H)}$, where $\tilde{\Delta}(H)$ is defined as follows. Given a dominating base F over an edge $e \in \mathsf{E}(H)$, we set

$$\tilde{d}_{\mathsf{F}}(e) = \max\{\tilde{d}_{\mathsf{f}}(e) : \mathsf{f} \in \mathsf{F}\},$$
$$\tilde{d}_{\mathsf{f}}(e) = \frac{d^H(e) - d^H_{\mathsf{f}}(e) + 1}{|e \setminus \mathsf{f}|} = \begin{cases} \frac{d^H(e\setminus \mathsf{f})}{|e\setminus \mathsf{f}|} & \mathsf{f} \neq \emptyset, \\ \frac{d^H(e)+1}{r} & \mathsf{f} = \emptyset. \end{cases}$$

(Compare (3.14).) Note that $\tilde{d}_{\mathsf{f}}(e)$ is the average degree of vertices in $e \setminus \mathsf{f}$. We define

$$\tilde{d}(e) = \min_{\mathsf{F}} \tilde{d}_{\mathsf{F}}(e), \qquad \tilde{\Delta}(H) = \max_{e\in\mathsf{E}(H)} \tilde{d}(e),$$

where the minimum is taken over all dominating bases F for e.

4. Proof of Theorem 2.15 (Counting lemma)

We will actually prove a more general version, involving a generalization of homomorphism counts that also includes *induced* homomorphism counts as a special case. We say that a pair $\mathcal{H} = (H, \xi)$ is a *signed hypergraph* if $H = (\mathsf{V}, \mathsf{E})$ is a hypergraph and $\xi : \mathsf{E} \to \{-1, +1\}$ is a labeling of the edges by signs. Recall from Section 1.6 that $H' = (\mathsf{V}', \mathsf{E}') \subseteq H$ if $\mathsf{V}' \subseteq \mathsf{V}$ and $\mathsf{E}' \subseteq \mathsf{E}$, and $H' \subset H$ if $\mathsf{V}' \subseteq \mathsf{V}$ and $\mathsf{E}' \subset \mathsf{E}$. We say that $\mathcal{H}' = (H', \xi') \subseteq \mathcal{H} = (H, \xi)$ (resp., $\mathcal{H}' = (H', \xi') \subset \mathcal{H}$) if $H' \subseteq H$ (resp., $H' \subset H$) and $\xi' = \xi|_{\mathsf{E}'}$. For a signed hypergraph $\mathcal{H} = (H, \xi)$, the signing induces two subgraphs of H given by $H_\pm$ with $\mathsf{V}(H_\pm) = \mathsf{V}(H)$ and $\mathsf{E}(H_\pm) = \xi^{-1}(\pm 1)$. We extend the definition of homomorphism counts to signed hypergraphs by defining for any $\mathcal{H}' \subseteq \mathcal{H}$ and $\underline{S} = (S^e)_{e\in\mathsf{E}(H)} \in \mathcal{S}^{\mathsf{E}(H)}$,

$$\hom(\mathcal{H}', \underline{S}) = \sum_{\phi:\mathsf{V}(H')\to[n]} \prod_{e\in\mathsf{E}(H'_+)} S^e(\phi_e) \prod_{e\in\mathsf{E}(H'_-)} \big(1 - S^e(\phi_e)\big). \tag{4.1}$$

For compactness, here and in the remainder of the section we write

$$\phi_v := \phi(v), \qquad \phi_e := \phi(e) = \{\phi_v\}_{v\in e}$$

and similarly $\phi_U := \phi(U)$ for general $U \subset \mathsf{V}$.

We can alternatively express this using the functional $\hom(H,\cdot)$ as follows: with ξ fixed, we denote

$$\widetilde{\underline{S}} = \widetilde{\underline{S}}_\xi = (\widetilde{S}^e)_{e\in\mathsf{E}(H)}, \qquad \widetilde{S}^e := \begin{cases} S^e & \xi(e) = +1, \\ J_{n,r} - S^e & \xi(e) = -1. \end{cases} \tag{4.2}$$

(Recall that $J_{n,r}$ is the tensor with entries 1 when all indices are distinct and 0 otherwise.) We have

$$\hom(\mathcal{H}, \underline{S}) = \hom(H, \widetilde{\underline{S}}). \tag{4.3}$$

Theorem 2.15 follows immediately from the next result upon taking the trivial labeling $\xi(e) \equiv 1$.

THEOREM 4.1 (Counting lemma for signed homomorphisms)
Let $p \in (0,1)$, and let $\mathcal{H} = (H,\xi)$ be a signed hypergraph as above. For each $e \in \mathsf{E}(H)$, let $\mathsf{F}(e)$ be an H-dominating base for e, and define a weighted base $\mathsf{B}(e) = (r, e, \iota_e, \mathsf{F}(e), d^{H_+}(e), \{d_{\mathsf{f}}^{H_+}(e)\}_{\mathsf{f}\in\mathsf{F}(e)})$. With base system $\mathbb{B} = \{\mathsf{B}(e)\}_{e\in\mathsf{E}(H)}$, let $\|\cdot\|_{\mathbb{B}}^*$ be the associated seminorm on $\mathcal{Z}_{n,r}$ as defined in (2.14). Let $\mathcal{C} \subset \mathcal{A}_{n,r}$ be a set of diameter at most εp under $\|\cdot\|_{\mathbb{B}}^*$ for some $\varepsilon \in (0,1]$, and assume further that there exist $Q_0 \in \mathrm{hull}(\mathcal{C})$ and $L \geq 1$ such that

$$\hom(\mathcal{H}', Q_0) \leq L n^{\mathsf{v}(H')} p^{\mathsf{e}(H'_+)} \quad \forall \mathcal{H}' = (H', \xi') \subset \mathcal{H}. \tag{4.4}$$

Then for all $\underline{P} = (P^e)_{e\in\mathsf{E}(H)}$ and $\underline{Q} = (Q^e)_{e\in\mathsf{E}(H)}$ with each $P^e, Q^e \in \mathrm{hull}(\mathcal{C})$,

$$\left|\hom(\mathcal{H}, \underline{P}) - \hom(\mathcal{H}, \underline{Q})\right| \lesssim_{\mathcal{H}} L\varepsilon n^{\mathsf{v}(H)} p^{\mathsf{e}(H_+)}. \tag{4.5}$$

Note that while the bases for $\mathsf{B}(e)$ are H-dominating, the weights are taken from the neighborhood structure in the subgraph H_+.

Proof
Fix $\mathcal{H}$ and $\mathcal{C}$ as in the statement of the lemma. We prove by induction on $m \leq \mathsf{e}(H)$ that for all $\mathcal{H}' = (H', \xi') \subseteq \mathcal{H}$ with $\mathsf{e}(H') \leq m$, all $\underline{A} = (A^e) \in \mathcal{C}^{\mathsf{E}(H')}$ and $\underline{Q} = (Q^e) \in \mathrm{hull}(\mathcal{C})^{\mathsf{E}(H')}$, we have

$$\left|\hom(\mathcal{H}', \underline{A}) - \hom(\mathcal{H}', \underline{Q})\right| \leq C(m,r) L\varepsilon n^{\mathsf{v}(H')} p^{\mathsf{e}(H'_+)} \tag{4.6}$$

for some $C(m,r) < \infty$. One can then replace $\underline{A}$ with $\underline{P}$ as in the theorem statement via the triangle inequality.

The base case $m = 0$ holds trivially. Assume now that (4.6) holds for all $\mathcal{H}' \subseteq \mathcal{H}$ with $\mathsf{e}(H') \leq m-1$. We fix $\mathcal{H}' = (H', \xi') \subseteq \mathcal{H}$ with $\mathsf{e}(H') = m$ and $\underline{A}$ and $\underline{Q}$ as above. For brevity, we write

$$\mathsf{V}' := \mathsf{V}(H'), \qquad \mathsf{E}' := \mathsf{E}(H'), \qquad \mathsf{E}'_{\pm} := \mathsf{E}(H'_{\pm}).$$

We first express $\hom(\mathcal{H}', \underline{Q})$ as a convex combination of homomorphism counts for Boolean tensors. Labeling the elements of $\mathcal{C}$ as B_j, $1 \le j \le |\mathcal{C}|$, for each $e \in \mathsf{E}(H)$ we express $Q^e = \sum_j c_j^e B_j$ for coefficients $c_j^e \in [0,1]$ with $\sum_j c_j^e = 1$. We have

$$\begin{aligned}
\hom(\mathcal{H}', \underline{Q}) &= \sum_{\phi:\mathsf{V}'\to[n]} \prod_{e\in\mathsf{E}'_+} Q^e(\phi_e) \prod_{e\in\mathsf{E}'_-} \big(1 - Q^e(\phi_e)\big) \\
&= \sum_{\phi:\mathsf{V}'\to[n]} \prod_{e\in\mathsf{E}'_+} \Big[\sum_j c_j^e B_j(\phi_e)\Big] \prod_{e\in\mathsf{E}'_-} \Big[\sum_j c_j^e \big(1 - B_j(\phi_e)\big)\Big] \\
&= \sum_{\phi:\mathsf{V}'\to[n]} \sum_{\boldsymbol{j}} \prod_{e\in\mathsf{E}'} c_{j_e}^e \prod_{e\in\mathsf{E}'_+} B_{j_e}(\phi_e) \prod_{e\in\mathsf{E}'_-} \big(1 - B_{j_e}(\phi_e)\big) \\
&= \sum_{\boldsymbol{j}} c_{\boldsymbol{j}} \hom(\mathcal{H}', \underline{B}_{\boldsymbol{j}}),
\end{aligned}$$

where sums over $\boldsymbol{j}$ run over all $\boldsymbol{j} = (j_e)_{e\in\mathsf{E}'} \in |\mathcal{C}|^{\mathsf{E}'}$, and we set

$$c_{\boldsymbol{j}} := \prod_{e\in\mathsf{E}'} c_{j_e}^e, \qquad \underline{B}_{\boldsymbol{j}} := (B_{j_e})_{e\in\mathsf{E}'}.$$

Now noting that $\sum_{\boldsymbol{j}} c_{\boldsymbol{j}} = 1$, we have

$$\begin{aligned}
\big|\hom(\mathcal{H}', \underline{A}) - \hom(\mathcal{H}', \underline{Q})\big| &= \Big|\sum_{\boldsymbol{j}} c_{\boldsymbol{j}} \big(\hom(\mathcal{H}', \underline{A}) - \hom(\mathcal{H}', \underline{B}_{\boldsymbol{j}})\big)\Big| \\
&\le \sum_{\boldsymbol{j}} c_{\boldsymbol{j}} \big|\hom(\mathcal{H}', \underline{A}) - \hom(\mathcal{H}', \underline{B}_{\boldsymbol{j}})\big|.
\end{aligned}$$

Thus, fixing collections $\underline{A} = (A^e)_{e\in\mathsf{E}'}$ and $\underline{B} = (B^e)_{e\in\mathsf{E}'}$ of tensors in $\mathcal{C}$, it suffices to show that

$$\big|\hom(\mathcal{H}', \underline{A}) - \hom(\mathcal{H}', \underline{B})\big| \le C(m,r) L \varepsilon n^{|\mathsf{V}'|} p^{|\mathsf{E}'_+|}. \tag{4.7}$$

Label the hyperedges of E' as $e_1, \dots, e_m$. Recalling the notation (4.2), we express the difference of homomorphism counts as a telescoping sum

$$\begin{aligned}
&\hom(\mathcal{H}', \underline{A}) - \hom(\mathcal{H}', \underline{B}) \\
&\quad = \hom(H', \widetilde{\underline{A}}) - \hom(H', \widetilde{\underline{B}})
\end{aligned}$$

$$= \sum_{\phi:\mathsf{V}'\to[n]} \sum_{k=1}^{m} \big[\widetilde{A}^{e_k}(\phi_{e_k}) - \widetilde{B}^{e_k}(\phi_{e_k})\big] \prod_{j<k} \widetilde{B}^{e_j}(\phi_{e_j}) \prod_{j>k} \widetilde{A}^{e_j}(\phi_{e_j})$$

$$= \sum_{k=1}^{m} \sum_{\phi:\mathsf{V}'\setminus e_k\to[n]} \widetilde{\Gamma}^{e_k}_{k,\phi} \prod_{\substack{e\in\mathsf{E}':\\ e\cap e_k=\emptyset}} \widetilde{Z}^{e}_{k}(\phi_e),$$

where

$$\mathcal{A}_{n,r} \ni \widetilde{Z}^{e}_{k} := \begin{cases} \widetilde{B}^{e} & \text{for } e = e_j \text{ with } j \le k, \\ \widetilde{A}^{e} & \text{for } e = e_j \text{ with } j > k, \end{cases}$$

$$\widetilde{\Gamma}^{e_k}_{k,\phi} := \sum_{\psi:e_k\to[n]} \big[\widetilde{A}^{e_k}(\psi_{e_k}) - \widetilde{B}^{e_k}(\psi_{e_k})\big] \prod_{e\in\partial H'(e_k)} \widetilde{Z}^{e}_{k}(\phi_{e\setminus e_k} \cup \psi_{e\cap e_k}),$$

and $\widetilde{Z}^{e}_{k}(\phi_{e\setminus e_k} \cup \psi_{e\cap e_k})$ is the value of the symmetric r-tensor $\widetilde{Z}^{e}_{k}$ evaluated at an arbitrary ordering of the set $\phi_{e\setminus e_k} \cup \psi_{e\cap e_k}$.

Now we recognize the expression

$$\prod_{e\in\partial H'(e_k)} \widetilde{Z}^{e}_{k}(\phi_{e\setminus e_k} \cup \psi_{e\cap e_k}) = \prod_{\mathsf{f}\in\mathsf{F}(e_k)} \prod_{e\in\partial_{\mathsf{f}} H'(e_k)} \widetilde{Z}^{e}_{k}(\phi_{e\setminus e_k} \cup \psi_{e\cap e_k})$$

$$=: \prod_{\mathsf{f}\in\mathsf{F}(e_k)} \tau_{\mathsf{f}}\big((\psi_v)_{v\in\mathsf{f}}\big) =: T_{e_k,\phi}\big((\psi_v)_{v\in e_k}\big) \tag{4.8}$$

as the output of a test tensor $T_{e_k,\phi} \in \mathcal{T}(e_k)$. Hence we can express

$$\widetilde{\Gamma}^{e_k}_{k,\phi} = \langle \widetilde{A}^{e_k} - \widetilde{B}^{e_k}, T_{e_k,\phi} \rangle.$$

Noting that $\widetilde{A}^{e_k} - \widetilde{B}^{e_k} = \pm(A^{e_k} - B^{e_k})$ for each k, we can apply the triangle inequality and our assumption on the diameter of $\mathcal{C}$ to bound

$$\big|\hom(\mathcal{H}', \underline{A}) - \hom(\mathcal{H}', \underline{B})\big|$$

$$\le \sum_{k=1}^{m} \sum_{\phi:\mathsf{V}'\setminus e_k\to[n]} \big|\langle A^{e_k} - B^{e_k}, T_{e_k,\phi}\rangle\big| \prod_{\substack{e\in\mathsf{E}':\\ e\cap e_k=\emptyset}} \widetilde{Z}^{e}_{k}(\phi_e)$$

$$\le \varepsilon p \sum_{k=1}^{m} \sum_{\phi:\mathsf{V}'\setminus e_k\to[n]} \|T_{e_k,\phi}\|_{\mathsf{B}(e_k)} \prod_{\substack{e\in\mathsf{E}':\\ e\cap e_k=\emptyset}} \widetilde{Z}^{e}_{k}(\phi_e). \tag{4.9}$$

Recalling our choice of weights for the weighted base $\mathsf{B}(e_k)$, by definition we have

$$\|T_{e_k,\phi}\|_{\mathsf{B}(e_k)} \le \|T_{e_k,\phi}\|_1 + \sum_{\mathsf{f}\in\mathsf{F}(e_k)} n^{r-|\mathsf{f}|}\, p^{d^{H}_{+}(e_k) - d^{H}_{+}_{\mathsf{f}}(e_k)} \|\tau_{\mathsf{f}}\|_1.$$

From (4.8),

$$\|T_{e_k,\phi}\|_1 = \sum_{\psi:e_k\to[n]} \prod_{e\in\partial^{H'}(e_k)} \widetilde{Z}_k^e(\phi_{e\setminus e_k} \cup \psi_{e\cap e_k})$$

and for $\|\tau_{\mathsf{f}}\|_1$ we have the same expression with $\partial_{\mathsf{f}}^{H'}(e_k)$ in place of $\partial^{H'}(e_k)$. Substituting these bounds in (4.9), we obtain

$$\begin{aligned}
&\big|\hom(\mathcal{H}', \underline{A}) - \hom(\mathcal{H}', \underline{B})\big| \\
&\quad\le \varepsilon p \sum_{k=1}^{m} \Big\{ \hom(H^{(k)}, \underline{\widetilde{Z}}_k) \\
&\qquad + \sum_{\mathsf{f}\in\mathsf{F}(e_k)} n^{r-|\mathsf{f}|}\, p^{d^{H_+}(e_k) - d_{\mathsf{f}}^{H_+}(e_k)} \hom(H^{(k,\mathsf{f})}, \underline{\widetilde{Z}}_k) \Big\},
\end{aligned} \tag{4.10}$$

where $H^{(k)} = (\mathsf{V}', \mathsf{E}' \setminus \{e_k\})$, and for $H^{(k,\mathsf{f})}$,

$$\mathsf{V}(H^{(k,\mathsf{f})}) = \mathsf{V}' \setminus (e_k \setminus \mathsf{f}), \qquad \mathsf{E}(H^{(k,\mathsf{f})}) = \big\{e \in \mathsf{E}' : e \cap (e_k \setminus \mathsf{f}) = \emptyset\big\}.$$

In particular,

$$\mathsf{v}(H^{(k,\mathsf{f})}) = |\mathsf{V}'| - r + |\mathsf{f}| \tag{4.11}$$

and

$$\begin{aligned}
\mathsf{e}(H_+^{(k,\mathsf{f})}) &= \big|\big\{e \in \mathsf{E}'_+ : e \cap (e_k \setminus \mathsf{f}) = \emptyset\big\}\big| \\
&= \big|\mathsf{E}'_+ \cap \{e_k\}^c \cap \partial^{H_+}(e_k) \cap \partial_{\mathsf{f}}^{H_+}(e_k)^c\big| \\
&= |\mathsf{E}'_+| - 1_{\xi(e_k)=1} - d^{H'_+}(e_k) + d_{\mathsf{f}}^{H'_+}(e_k) \\
&\ge |\mathsf{E}'_+| - 1_{\xi(e_k)=1} - d^{H_+}(e_k) + d_{\mathsf{f}}^{H_+}(e_k).
\end{aligned} \tag{4.12}$$

By restricting ξ to the edge sets of $H^{(k)}$ and $H^{(k,\mathsf{f})}$, we obtain signed hypergraphs $\mathcal{H}^{(k)} \subset \mathcal{H}'$ and $\mathcal{H}^{(k,\mathsf{f})} \subset \mathcal{H}'$ for each $k \in [m]$ and $\mathsf{f} \in \mathsf{F}(e_k)$. For any $\mathcal{H}'' = (H'', \xi'')$ in this collection of signed hypergraphs we have $\mathsf{e}(H'') \le m-1$, and by the induction hypothesis and the assumption (4.4), for any $\underline{Q} \in \mathrm{hull}(\mathcal{C})^{\mathsf{E}'}$,

$$\begin{aligned}
\hom(H'', \underline{Q}) &= \hom(\mathcal{H}'', \underline{Q}) \le \hom(\mathcal{H}'', Q_0) + \big|\hom(\mathcal{H}'', \underline{Q}) - \hom(\mathcal{H}'', A_0)\big| \\
&\le \big(1 + C(m-1, r)\big) L n^{\mathsf{v}(H'')} p^{\mathsf{e}(H''_+)}
\end{aligned}$$

(recalling $\varepsilon \le 1$). Applying this for each $\mathcal{H}^{(k)}$ and $\mathcal{H}^{(k,\mathsf{f})}$ with $\underline{Q} = \underline{Z}_k$ and combining with (4.11) and (4.12), we obtain

$$\hom(\mathscr{H}^{(k)}, \underline{\widetilde{Z}}_k) \le \big(1 + C(m-1,r)\big) L n^{|\mathsf{V}'|} p^{\mathsf{e}(H'_+)-1},$$

$$\hom(\mathscr{H}^{(k,\mathsf{f})}, \underline{\widetilde{Z}}_k) \le \big(1 + C(m-1,r)\big) L n^{|\mathsf{V}'|-r+|\mathsf{f}|} p^{\mathsf{e}(H'_+)-1-d^{H_+}(e_k)+d_{\mathsf{f}}^{H_+}(e_k)}.$$

Substituting these bounds into (4.10), we obtain (4.7) upon taking $C(m,r) := m2^r(1 + C(m-1,r))$. This completes the induction step to conclude the proof of Theorem 4.1. □

5. Proof of Theorem 2.13 (Decomposition lemma)

Throughout this section, we write $\mathcal{T} := \mathcal{T}_{\mathsf{B}}$. For $A \in \mathscr{A}_n$, we denote the centered tensor

$$\bar{A} = A - \mathbb{E}\boldsymbol{A} = A - pJ_{n,r}.$$

We refer the reader to Section 1.6 for our notational conventions for tensors.

LEMMA 5.1

Let $k \ge 1$ and $T_1, \dots, T_k \in \mathcal{T}$. For $1 \le i \le k$, let W_i be the span of $\{T_1, \dots, T_i\}$ and set

$$\widehat{T}_i := P_{W_{i-1}^{\perp}}(T_i) \tag{5.1}$$

(with $\widehat{T}_1 = T_1$). We have

$$\mathbb{P}\Big(\bigwedge_{i\in[k]} |\langle \bar{\boldsymbol{A}}, \widehat{T}_i\rangle| \ge \varepsilon p \|T_i\|_{\mathsf{B}}\Big) \le 2^k \exp\Big(-c(r)\varepsilon^2 p^2 \big(1 \vee \log(1/p)\big) \sum_{i=1}^k \|T_i\|_{\mathsf{B}}\Big)$$

for some $c(r) > 0$ depending only on r.

Proof

By the union bound,

$$\mathbb{P}\Big(\bigwedge_{i\in[k]} |\langle \bar{\boldsymbol{A}}, \widehat{T}_i\rangle| \ge \varepsilon p \|T_i\|_{\mathsf{B}}\Big) \le 2^k \mathbb{P}\Big(\bigwedge_{i\in[k]} \langle \bar{\boldsymbol{A}}, \sigma_i \widehat{T}_i\rangle \ge \varepsilon p \|T_i\|_{\mathsf{B}}\Big),$$

where $\sigma_i \in \{+1, -1\}$. Fix a choice of $\sigma_1, \dots, \sigma_k$, and let $\widetilde{T}_i = \sigma_i \widehat{T}_i$. Then

$$\mathbb{P}\Big(\bigwedge_{i\in[k]} \langle \bar{\boldsymbol{A}}, \sigma_i \widehat{T}_i\rangle \ge \varepsilon p \|T_i\|_{\mathsf{B}}\Big) \le \mathbb{P}\Big(\Big\langle \bar{\boldsymbol{A}}, \sum_{i=1}^k \widetilde{T}_i\Big\rangle \ge \varepsilon p \sum_{i=1}^k \|T_i\|_{\mathsf{B}}\Big).$$

Since $\widetilde{T}_1, \dots, \widetilde{T}_k$ are orthogonal, we have

$$\Big\|\sum_{i=1}^k \widetilde{T}_i\Big\|_2^2 = \sum_{i=1}^k \|\widetilde{T}_i\|_2^2 \le \sum_{i=1}^k \|T_i\|_2^2 = \sum_{i=1}^k \|T_i\|_1, \tag{5.2}$$

where the last equality uses that the T_i are Boolean tensors.

Let $\widetilde{T} = \sum_{i=1}^{k} \widetilde{T}_i$. Then for any $\lambda > 0$,

$$\mathbb{P}\Big(\Big\langle \bar{\boldsymbol{A}}, \sum_{i=1}^{k} \widetilde{T}_i \Big\rangle \geq \varepsilon p \sum_{i=1}^{k} \|T_i\|_{\square} \Big) \leq \exp\Big(-\lambda \varepsilon p \sum_{i=1}^{k} \|T_i\|_{\square}\Big) \mathbb{E} \exp\big(\lambda \langle \bar{\boldsymbol{A}}, \widetilde{T} \rangle\big).$$

Recall that the entries of $\bar{\boldsymbol{A}}$ with distinct coordinates are independent centered Bernoulli(p) random variables, up to the symmetry constraint. Let $\bar{\boldsymbol{A}}' : \{n\}^r / S_r \to \mathbb{R}$ be the independent entries of $\bar{\boldsymbol{A}}$. Notice that

$$\langle \bar{\boldsymbol{A}}, \widetilde{T} \rangle = \langle \bar{\boldsymbol{A}}', \widetilde{T}' \rangle$$

for some tensor $\widetilde{T}'$ in which each coordinate is a sum of at most $r!$ entries of $\widetilde{T}$. We thus have $\|\widetilde{T}'\|_2^2 \leq r! \|\widetilde{T}\|_2^2$. Hence,

$$\mathbb{E} \exp\big(\lambda \langle \bar{\boldsymbol{A}}, \widetilde{T} \rangle\big) = \mathbb{E} \exp\big(\lambda \langle \bar{\boldsymbol{A}}', \widetilde{T}' \rangle\big) = \prod_{\mathbf{i} \in [n]^r / S} \big(p e^{\lambda(1-p)\widetilde{T}'(\mathbf{i})} + (1-p) e^{-\lambda p \widetilde{T}'(\mathbf{i})}\big).$$

We claim that

$$p^{\lambda(1-p)x} + (1-p) e^{-\lambda p x} \leq e^{\lambda^2 x^2 / \log(1/p)}.$$

Indeed, we have

$$p e^{\lambda(1-p)x} + (1-p) e^{-\lambda p x} \leq \exp\big(-\lambda p x + p[e^{\lambda x} - 1]\big) \leq \exp\big(\lambda^2 x^2 / \log(1/p)\big)$$

assuming that $|\lambda x| \leq \log(1/p)$, since for $|z| \leq \log(1/p)$, we have $\exp(z) \leq 1 + z + z^2/(p \log(1/p)^2)$ by monotonicity of the function $z \mapsto \frac{\exp(z) - 1 - z}{z^2}$. Otherwise, $|\lambda x| > \log(1/p)$ and we have

$$\begin{aligned} p e^{\lambda(1-p)x} + (1-p) e^{-\lambda p x} &= e^{\lambda(1-p)x + \log p} + (1-p) e^{-\lambda p x} \\ &\leq \exp\big(\lambda^2 x^2 / \log(1/p)\big). \end{aligned}$$

Thus,

$$\mathbb{E} \exp\big(\lambda \langle \bar{\boldsymbol{A}}, \widetilde{T} \rangle\big) \leq \exp\big(\lambda^2 \|\widetilde{T}'\|_2^2 / \log(1/p)\big) \leq \exp\big(c_1(r) \lambda^2 \|\widetilde{T}\|_2^2 / \log(1/p)\big).$$

By choosing $\lambda = c_2(r) \varepsilon p \log(1/p) \sum_{i=1}^{k} \|T_i\|_{\square} / \|\widetilde{T}\|_2^2$, we obtain

$$\begin{aligned} &\mathbb{P}\Big(\Big\langle \bar{\boldsymbol{A}}, \sum_{i=1}^{k} \widetilde{T}_i \Big\rangle \geq \varepsilon p \sum_{i=1}^{k} \|T_i\|_{\square} \Big) \\ &\quad \leq \exp\Big(-c_3(r) \varepsilon^2 p^2 \log(1/p) \Big(\sum_{i=1}^{k} \|T_i\|_{\square}\Big)^2 \Big/ \|\tilde{T}\|_2^2\Big) \\ &\quad \leq \exp\Big(-c(r) \varepsilon^2 p^2 \log(1/p) \sum_{i=1}^{k} \|T_i\|_{\square}\Big), \end{aligned}$$

using (5.2) and $\sum_{i=1}^{k} \|T_i\|_1 \le \sum_{i=1}^{k} \|T_i\|_{\mathsf{B}}$. Moreover, the same bound with $\log(1/p)$ replaced by 1 follows from Hoeffding's inequality. □

We establish Theorem 2.13 by the following iterative procedure. We initialize $R_0 = \bar{A}$. If $\|R_0\|_{\mathsf{B}}^* \le \varepsilon p$, then the claim follows with $k = 0$. Otherwise, we proceed to step $k = 1$. At step $k \ge 1$, having obtained $R_0, \ldots, R_{k-1}$ and $T_1, \ldots, T_{k-1}$, if $\|R_{k-1}\|_{\mathsf{B}}^* > \varepsilon p$, then there exists $T_k \in \mathcal{T}$ so that $|\langle R_{k-1}, T_k\rangle| > \varepsilon p \|T_k\|_{\mathsf{B}}$. Taking such a T_k, we set

$$R_k = R_{k-1} - P_{\mathrm{span}(\widehat{T}_k)}(R_{k-1}) = P_{W_k^{\perp}}(\bar{A}),$$

where for brevity we denote the subspace $W_k := \mathrm{span}(T_1, \ldots, T_k)$. We stop the process at step k if either

$$\|R_k\|_{\mathsf{B}}^* \le \varepsilon p \qquad \text{or} \qquad \sum_{i=1}^{k} \|T_i\|_{\mathsf{B}} > \kappa \varepsilon^{-2} n^r p^{-2}, \tag{5.3}$$

and otherwise proceed to step $k + 1$. Note that the process must stop at step k for some

$$k \le k_\star := \lfloor 1 + \kappa \varepsilon^{-2} p^{-d_\star - 2} \rfloor. \tag{5.4}$$

Indeed, if the process has not stopped after step $k - 1$ for some $k \ge 1$, then $\sum_{i=1}^{k-1} \|T_i\|_{\mathsf{B}} \le \kappa \varepsilon^{-2} n^r p^{-2}$, while on the other hand $\|T_i\|_{\mathsf{B}} \ge \|T_i\|_{\emptyset} = n^r p^{d_\star}$ for each $1 \le i \le k - 1$, and (5.4) follows by combining these bounds.

We take $\sum_{i=1}^{k} \alpha_i T_i$ to be the expansion of $P_{W_k}(\bar{A})$ in the basis $\{T_1, \ldots, T_k\}$. Note that for each $1 \le j \le k$, since R_{j-1} is orthogonal to $T_1, \ldots, T_{j-1}$,

$$\varepsilon p \|T_j\|_{\mathsf{B}} < \left|\langle R_{j-1}, T_j\rangle\right| = \left|\langle R_{j-1}, \widehat{T}_j\rangle\right| \le \|R_{j-1}\|_2 \|\widehat{T}_j\|_2 \le \|A\|_2 \|\widehat{T}_j\|_2$$

(recalling the notation (5.1)), and so the distance of T_j to the span of $\{T_1, \ldots, T_{j-1}\}$ is

$$\|\widehat{T}_j\|_2 \ge \frac{\varepsilon p \|T_j\|_{\mathsf{B}}}{\|A\|_2} \ge \frac{\varepsilon p n^r p^{d_\star}}{n^{r/2}} = \varepsilon p^{1+d_\star} n^{r/2},$$

as claimed.

If the process stops at some k for which $\sum_{i=1}^{k} \|T_i\|_{\mathsf{B}} \le \kappa \varepsilon^{-2} n^r p^{-2}$, then the first condition in (5.3) holds, that is,

$$\left\|\bar{A} - P_{W_k}(\bar{A})\right\|_{\mathsf{B}}^* \le \varepsilon p,$$

and we obtain the claim. We take $\mathcal{E}_\star(\kappa, \varepsilon)$ to be the set of $A \in \mathcal{A}_{n,r}$ for which the process runs until the second condition in (5.3) holds for some $k \le k_\star$.

Thus, it only remains to bound the measure of $\mathcal{E}_\star(\kappa,\varepsilon)$. For the case that the process ends at step $k=1$ we obtained the desired probability bound from the second bound in (5.3) and Lemma 5.1, so we may henceforth assume that $k\geq 2$. In particular, from (5.4) it follows that $\kappa\varepsilon^{-2}p^{-d_\star-2}\geq 1$ in this case. Denoting the event in Lemma 5.1 by $\mathcal{E}(T_1,\dots,T_k)$, we have

$$\mathcal{E}_\star(\kappa,\varepsilon)\subseteq \bigcup_{\substack{T_1,\dots,T_k:\\ \sum_{i=1}^k \|T_i\|_{\mathsf{B}}>\kappa\varepsilon^{-2}n^r p^{-2}}} \mathcal{E}(T_1,\dots,T_k). \tag{5.5}$$

By Lemma 5.1, for each fixed sequence $T_1,\dots,T_k$,

$$\mathbb{P}\big(\mathcal{E}(T_1,\dots,T_k)\big)\leq 2^k\exp\Big(-c(r)\varepsilon^2p^2\big(1\vee\log(1/p)\big)\sum_{i=1}^k\|T_i\|_{\mathsf{B}}\Big). \tag{5.6}$$

We break up the union on the right-hand side of (5.5) into dyadic ranges for $\sum_{i=1}^k\|T_i\|_{\mathsf{B}}$. For each $j\geq 0$, let

$$\mathcal{E}_{\star j}(\kappa,\varepsilon):=\bigcup_{\substack{T_1,\dots,T_k:\\ \sum_{i=1}^k\|T_i\|_{\mathsf{B}}\in I_j}}\mathcal{E}(T_1,\dots,T_k),$$

where $I_j:=\kappa\varepsilon^{-2}n^rp^{-2}\cdot[2^j,2^{j+1})$. Writing $T_i=\prod_{\mathsf{f}\in\mathsf{F}}\tau_{\mathsf{f}}^{(i)}\circ\pi_{\mathsf{f}}$ as in (2.10) we have that for all $\mathsf{f}\in\mathsf{F}$,

$$\|\tau_{\mathsf{f}}^{(i)}\|_1\leq\|T_i\|_{\mathsf{B}}/(n^{r-|\mathsf{f}|}p^{d_\star-d_{\mathsf{f}}}).$$

The number of choices for the Boolean tensor $\tau_{\mathsf{f}}^{(i)}$ given $\|T_i\|_{\mathsf{B}}$ is thus at most

$$n^{|\mathsf{f}|\|\tau_{\mathsf{f}}^{(i)}\|_1}\leq\exp\Big(\frac{|\mathsf{f}|\cdot\|T_i\|_{\mathsf{B}}(\log n)}{n^{r-|\mathsf{f}|}p^{d_\star-d_{\mathsf{f}}}}\Big), \tag{5.7}$$

and so the number of choices for T_i given $\|T_i\|_{\mathsf{B}}$ is at most

$$\exp\big(r2^r(\log n)\cdot\|T_i\|_{\mathsf{B}}\cdot\max_{\mathsf{f}\in\mathsf{F}}\{n^{|\mathsf{f}|-r}p^{d_{\mathsf{f}}-d_\star}\}\big).$$

Since each $\|T_i\|_{\mathsf{B}}$ can take at most $O_r(b)$ different values in an interval $[a,b]$, the total number of choices of $T_1,\dots,T_k$ with $\sum_{i=1}^k\|T_i\|_{\mathsf{B}}\in I_j$ is at most

$$\sum_{z_1+\cdots+z_k\in I_j}\exp\Big(r2^r(\log n)\cdot\Big(\sum_{i=1}^k z_i\Big)\cdot\max_{\mathsf{f}\in\mathsf{F}}\{n^{|\mathsf{f}|-r}p^{d_{\mathsf{f}}-d_\star}\}\Big)$$
$$\leq\exp\big(2^{j+1+2r}\kappa\varepsilon^{-2}(\log n)p^{-d_\star-2}\max_{\mathsf{f}\in\mathsf{F}}\{n^{|\mathsf{f}|}p^{d_{\mathsf{f}}}\}$$

$$+ O_r\big(k \log(2^{j+1}\kappa\varepsilon^{-2}n^r p^{-2})\big)\big)$$
$$= \exp\big(O_r(2^j)\cdot\kappa\varepsilon^{-2}(\log n)p^{-d_\star-2}\cdot\max_{\mathsf{f}\in\mathsf{F}}\{n^{|\mathsf{f}|}p^{d_\mathsf{f}}\}\big),$$

where we used that

$$\max_{\mathsf{f}\in\mathsf{F}}\{n^{|\mathsf{f}|}p^{d_\mathsf{f}}\} \geq n^{|\emptyset|}p^{d_\emptyset} = 1$$

along with (5.4) to absorb the errors depending on k (recall that we reduced to the case $\kappa\varepsilon^{-2}p^{-d_\star-2} \geq 1$, and note that $\kappa\varepsilon^{-2} = O(n)$ from our assumptions). Combining with (5.6), our assumption (2.16), and taking the constant $C_0 = C_0(r)$ there sufficiently large, we obtain

$$\begin{aligned}
&\mathbb{P}\big(\mathcal{E}_{\star j}(\kappa,\varepsilon)\big)\\
&\quad\leq \exp\big(O_r(2^j)\cdot\kappa\varepsilon^{-2}(\log n)p^{-d_\star-2}\cdot\max_{\mathsf{f}\in\mathsf{F}}\{n^{|\mathsf{f}|}p^{d_\mathsf{f}}\} - c(r)2^j\kappa n^r\big(1\vee\log(1/p)\big)\big)\\
&\quad\leq \exp\big(-c(r)2^j\kappa n^r\big(1\vee\log(1/p)\big)\big)
\end{aligned}$$

for a modified constant $c(r) > 0$. Summing the above bound over j and combining with (5.5) and the union bound, this completes the proof of Theorem 2.13. □

6. Proof of Theorem 3.1 (Quantitative LDP)

6.1. *Proof of the LDP upper bound*

In this section we prove Theorem 3.1(a). We use the following nonasymptotic LDP upper bound for convex sets, which holds in wide generality, and is a simple consequence of the minimax theorem (see [36, Exercise 4.5.5]).

LEMMA 6.1

For a Borel probability measure μ on a topological vector space $\mathcal{V}$, and any convex, compact subset $\mathcal{B} \subset \mathcal{V}$, we have

$$\mu(\mathcal{B}) \leq \exp\big(-\inf_{Q\in\mathcal{B}}\Lambda^*(Q)\big), \tag{6.1}$$

where $\Lambda^* : \mathcal{V} \to \mathbb{R}_+$ is the convex dual of the log-moment generating function $\Lambda(\lambda) = \log\int_{\mathcal{V}} e^{\lambda(Q)}\,d\mu(Q)$ on the dual vector space $\mathcal{V}^*$.

For the case of the Erdős–Rényi measure μ_p on the vector space $\mathcal{S}_{n,r}$ of real symmetric r-tensors with zero diagonals, one checks that $\Lambda^*(Q) = \mathrm{I}_p(Q)$.

We commence with the proof of Theorem 3.1(a). For $t \geq 0$ and sequences $\boldsymbol{T} = (T_1,\ldots,T_k) \in \mathcal{T}(e)^k$ and $\boldsymbol{\lambda} = (\lambda_1,\ldots,\lambda_k) \in \mathbb{R}^k$, we let $\mathcal{K}_e(\boldsymbol{T},\boldsymbol{\lambda};t)$ be the convex hull of all $A \in \mathcal{A}_{n,r}$ such that

$$\Big\| A - \mathbb{E}\boldsymbol{A} - \sum_{i=1}^{k} \lambda_i \widehat{T}_i \Big\|_{\mathsf{B}(e)}^{*} \le t, \tag{6.2}$$

where we recall from (5.1) the notation $(\widehat{T}_1, \dots, \widehat{T}_k)$ for the associated orthogonal sequence. For each $e \in \mathsf{E}$, let $\mathcal{J}_e$ be the collection of all sets of the form $\mathcal{K}_e(\boldsymbol{T}, \boldsymbol{\lambda}; 2\varepsilon p)$ for some $1 \le k \le \lfloor 1 + \kappa\varepsilon^{-2} p^{-d_\star(e)-2} \rfloor$, some $\boldsymbol{T} = (T_1, \dots, T_k) \in \mathcal{T}(e)^k$, and some $\boldsymbol{\lambda}$ in the scaled integer lattice $\Lambda^k := (\varepsilon p^{1+d_\star(e)/2}/k) \cdot \mathbb{Z}^k$ such that

$$\sum_{i=1}^{k} \|T_i\|_{\mathsf{B}(e)} \le \kappa\varepsilon^{-2} n^r p^{-2} \qquad \text{and} \qquad \|\boldsymbol{\lambda}\|_\infty \le p^{-1-d_\star(e)}\varepsilon^{-1}. \tag{6.3}$$

We claim that for each $e \in \mathsf{E}$,

$$\mathbb{P}\Big\{ \boldsymbol{A} \notin \bigcup_{\mathcal{K} \in \mathcal{J}_e} \mathcal{K} \Big\} \le \exp\big(-c_0 \kappa n^r \log(1/p)\big) \tag{6.4}$$

with $c_0 > 0$ as in Theorem 2.13. Indeed, it suffices to show that $\mathcal{J}_e$ covers the complement in $\mathcal{A}_{n,r}$ of the exceptional set $\mathcal{E}_{\star e}(\kappa, \varepsilon)$ provided by the application of Theorem 2.13 with weighted base $\mathsf{B}(e)$. To that end, fix an arbitrary $A \in \mathcal{A}_{n,r} \setminus \mathcal{E}_{\star e}(\kappa, \varepsilon)$. From Theorem 2.13 we have that A satisfies (6.2) with $t = \varepsilon p$ for some $\boldsymbol{T} \in \mathcal{T}(e)^k$ and $\boldsymbol{\lambda} \in \mathbb{R}^k$, with $\|\widehat{T}_j\|_2 \ge \varepsilon p^{1+d_\star(e)} n^{r/2}$ for each $1 \le j \le k$, where we henceforth write $d_\star(e)$, $d_{\mathrm{f}}(e)$ for the weights associated to the weighted base $\mathsf{B}(e)$. It follows from the Cauchy–Schwarz inequality that

$$|\lambda_j| = \frac{|\langle P_{W_j}(\bar{A}), \widehat{T}_j \rangle|}{\|\widehat{T}_j\|_2^2} \le \frac{\|\bar{A}\|_2}{\|\widehat{T}_j\|_2} \le \varepsilon^{-1} p^{-1-d_\star(e)},$$

so $\|\boldsymbol{\lambda}\|_\infty \le \varepsilon^{-1} p^{-1-d_\star(e)}$. Now let $\boldsymbol{\lambda}' \in \Lambda^k$ be as in (6.3) with $\|\boldsymbol{\lambda} - \boldsymbol{\lambda}'\|_\infty \le \varepsilon p^{1+d_\star(e)/2}/k$. By an application of the triangle inequality for the $\|\cdot\|_{\mathsf{B}(e)}^{*}$-seminorm, we only need to show that

$$\|\widehat{T}_i\|_{\mathsf{B}(e)}^{*} \le p^{-d_\star(e)/2} \tag{6.5}$$

for each $1 \le i \le k$. For this, note that $\|\widehat{T}_i\|_2 \le \|T_i\|_2 \le n^{r/2}$ since T_i is Boolean. Now for any $Z \in \mathcal{Z}_{n,r}$ with $\|Z\|_2 \le n^{r/2}$,

$$\|Z\|_{\mathsf{B}(e)}^{*} = \sup_{T \in \mathcal{T}(e)} \frac{|\langle Z, T \rangle|}{\|T\|_{\mathsf{B}(e)}} \le \frac{\|Z\|_2 \|T\|_2}{\|T\|_1^{1/2} (n^r p^{d_\star(e)})^{1/2}} = \frac{\|Z\|_2}{(n^r p^{d_\star(e)})^{1/2}} \le p^{-d_\star(e)/2},$$

where in the second equality we used that $\|T\|_2^2 = \|T\|_1$ for Boolean T. Thus we obtain (6.5) and hence (6.4), as desired.

Now set

$$\mathcal{I}'_{\mathsf{E}} = \Big\{ \bigcap_{e \in \mathsf{E}} \mathcal{K}_e : \mathcal{K}_e \in \mathcal{I}_e \text{ for each } e \in \mathsf{E} \Big\},$$

and let $\mathcal{I}_{\mathsf{E}}$ be obtained by replacing each $\mathcal{K} \in \mathcal{I}'_H$ with the convex hull of $\mathcal{K} \cap \mathcal{A}_{n,r}$. We claim that

$$\log |\mathcal{I}_{\mathsf{E}}| \lesssim_{\mathbb{B}} \kappa n^r \varepsilon^{-2} w_{n,p}^{-1} \log n. \tag{6.6}$$

Fixing $e \in \mathsf{E}$, it suffices to prove that the claimed bound holds for $\log |\mathcal{I}_e|$ (up to modification of the constant by a factor $|\mathsf{E}|$). First, recalling the bound (5.7), the number of $T \in \mathcal{T}(e)$ with a given value of $\|T\|_{\mathsf{B}(e)}$ is at most

$$\prod_{\mathsf{f} \in \mathsf{F}(e)} n^{|\mathsf{f}| \|T\|_{\mathsf{B}(e)} / (n^{r-|\mathsf{f}|} p^{d_\star(e) - d_{\mathsf{f}}(e)})},$$

so the total number of choices for $\boldsymbol{T}$ as in (6.3) is at most

$$\begin{aligned} &\exp\Big((\log n) \sum_{i=1}^{k} \sum_{\mathsf{f} \in \mathsf{F}(e)} \frac{|\mathsf{f}| \|T_i\|_{\mathsf{B}(e)}}{n^{r-|\mathsf{f}|} p^{d_\star(e) - d_{\mathsf{f}}(e)}} \Big) \\ &\qquad \le \exp\big(O_r(1) w_{n,p}(\mathsf{B}(e)) \big)^{-1} \kappa \varepsilon^{-2} n^r \log n \big). \end{aligned} \tag{6.7}$$

The number of choices for $k, \lambda_1, \dots, \lambda_k$ and $\|T_1\|_{\mathsf{B}(e)}, \dots, \|T_k\|_{\mathsf{B}(e)}$ is

$$\begin{aligned} &\sum_{k \le 1 + \kappa \varepsilon^{-2} p^{-d_\star(e)-2}} \Big(\frac{2}{\varepsilon^2 p^{2 + \frac{3}{2} d_\star(e)}} \Big)^k O_r(\kappa \varepsilon^{-2} n^r p^{-2})^k \\ &\qquad = n^{O_r(1)} n^{O_r(\kappa \varepsilon^{-2} p^{-d_\star(e)-2})}, \end{aligned} \tag{6.8}$$

where we noted that the bases of the exponentials in k are all $n^{O_r(1)}$ by our assumptions on n, p, κ, and ε. Now since $w_{n,p}(\mathsf{B}(e)) \le n^{r-|\emptyset|} p^{d_\star(e) - d_\emptyset + 2} = n^r p^{d_\star(e)+2}$, we see that the second factor in (6.8) is dominated by the right-hand side of (6.7). We thus obtained the claimed bound on $|\mathcal{I}_e|$, establishing (6.6).

Fix $\mathcal{E} \subseteq \mathcal{A}_{n,r}$. We claim that for any $\mathcal{K} \in \mathcal{I}_{\mathsf{E}}$,

$$\mathcal{K} \cap \mathcal{E} \ne \emptyset \Longrightarrow \mathcal{K} \subseteq (\mathcal{E})_{\mathbb{B}, 4\varepsilon p}. \tag{6.9}$$

Indeed, fix arbitrary $\mathcal{K} \in \mathcal{I}_{\mathsf{E}}$ with $\mathcal{K} \cap \mathcal{E} \ne \emptyset$. It suffices to show that for any fixed $A_1, A_2 \in \mathcal{K}$, we have

$$\|A_1 - A_2\|^*_{\mathsf{B}(e)} \le 4\varepsilon p \quad \forall e \in \mathsf{E}.$$

But this is immediate from the definitions: we have $\mathcal{K} = \bigcap_{e \in \mathsf{E}} \mathcal{K}_e$ for some choices of $\mathcal{K}_e \in \mathcal{I}_e$, and each $\mathcal{K}_e$ is contained in the $2\varepsilon p$-neighborhood of some $A'_e \in \mathcal{A}_{n,r}$ under $\|\cdot\|^*_{\mathsf{B}(e)}$, so the above bound follows by the triangle inequality.

Now we are ready to conclude. For $\mathcal{F} \subset \mathcal{Q}_{n,r}$, we abbreviate

$$\mathrm{I}_p(\mathcal{F}) := \inf\{\mathrm{I}_p(Q) : Q \in \mathcal{F}\}. \tag{6.10}$$

Applying the union bound and (6.4), we have

$$\mathbb{P}(\boldsymbol{A} \in \mathcal{E}) \le |\mathsf{E}|\, p^{c_0 \kappa n^r} + \sum_{\mathcal{K} \in \mathcal{J}_{\mathsf{E}} : \mathcal{K} \cap \mathcal{E} \neq \emptyset} \mathbb{P}(\boldsymbol{A} \in \mathcal{K}).$$

For the latter term we apply (6.1) to bound

$$\begin{aligned}
&\sum_{\mathcal{K} \in \mathcal{J}_{\mathsf{E}} : \mathcal{K} \cap \mathcal{E} \neq \emptyset} \mathbb{P}(\boldsymbol{A} \in \mathcal{K}) \\
&\quad \le \sum_{\mathcal{K} \in \mathcal{J}_{\mathsf{E}} : \mathcal{K} \cap \mathcal{E} \neq \emptyset} \exp\bigl(-\mathrm{I}_p(\mathcal{K})\bigr) \le |\mathcal{J}_{\mathsf{E}}| \max_{\mathcal{K} \in \mathcal{J}_{\mathsf{E}} : \mathcal{K} \cap \mathcal{E} \neq \emptyset} \exp\bigl(-\mathrm{I}_p(\mathcal{K})\bigr) \\
&\quad = |\mathcal{J}_{\mathsf{E}}| \exp\Bigl(-\mathrm{I}_p\Bigl(\bigcup_{\mathcal{K} \in \mathcal{J}_{\mathsf{E}} : \mathcal{K} \cap \mathcal{E} \neq \emptyset} \mathcal{K}\Bigr)\Bigr) \le |\mathcal{J}_{\mathsf{E}}| \exp\bigl(-\mathrm{I}_p\bigl((\mathcal{E})_{\mathbb{B},4\varepsilon p}\bigr)\bigr),
\end{aligned}$$

where in the final line we used (6.9). Applying (6.6), the claim now follows by using $\log(a+b) \le \max(\log(2a), \log(2b))$.

6.2. Proof of the LDP lower bound (3.8)
The claim quickly follows from the next lemma.

LEMMA 6.2
Fix a collection $\mathbb{B} = \{\mathsf{B}(e)\}_{e \in \mathsf{E}}$ of weighted bases, let $p \in (0,1)$, $\varepsilon > 0$, and assume that

$$\min_{e \in \mathsf{E}} w_{n,p}\bigl(\mathsf{B}(e)\bigr) \ge C(r) \varepsilon^{-2} \log n \tag{6.11}$$

for a sufficiently large constant $C(r) > 0$. Then for any $Q \in \mathcal{Q}_{n,r}$,

$$\log \mu_p\bigl(\mathcal{U}_{\mathbb{B}}(Q, \varepsilon p)\bigr) \ge -\mathrm{I}_p(Q) - O\bigl(1 + n^{r/2} \bigl|\log\bigl(p \wedge (1-p)\bigr)\bigr|\bigr).$$

Indeed, for any $\mathcal{E} \subseteq \mathcal{A}_{n,r}$ and any $Q \in (\mathcal{E})^{\circ}_{\mathbb{B},\varepsilon p}$ we have $\mathbb{P}(\boldsymbol{A} \in \mathcal{E}) \ge \mu_p(\mathcal{U}_{\mathbb{B}}(Q, \varepsilon p))$ by monotonicity, and the claim follows from the above lemma and taking the supremum over such Q.

It only remains to establish Lemma 6.2. To this end, we use the following two lemmas, starting with a complementary lower bound for Lemma 6.1.

LEMMA 6.3
For $d \ge 1$ and $p \in (0,1)$, let μ_p be the product Bernoulli(p) measure on $\{0,1\}^d$, and let ν be any other product measure on $\{0,1\}^d$. For any $\mathcal{E} \subseteq \{0,1\}^d$,

$$\log \mu_p(\mathcal{E}) \geq -\mathrm{D}(\nu\|\mu_p) + \log \nu(\mathcal{E}) - O\Big(\frac{d^{1/2}|\log(p \wedge (1-p))|}{\nu(\mathcal{E})^{1/2}}\Big).$$

Proof
While [31, Lemma 6.3] is stated in terms of two adjacency matrices of simple graphs of $d = \binom{n}{2}$ edges, one drawn from μ_p for some $p \in (0, \frac{1}{2}]$ and the other from some product measure ν on $\{0,1\}^d$, this lemma and its proof apply to any value of d and to any such product measures of Bernoulli variables. Further, with $p \leq \frac{1}{2}$ used only for the elementary bound of [31, (6.7)], upon changing p there to $(1-p)$, the same proof applies also for $p \in (\frac{1}{2}, 1)$. □

LEMMA 6.4
Fix a weighted base $\mathsf{B} = (r, e, \iota, \mathsf{F}, d_\star, \{d_\mathsf{f}\}_{\mathsf{f}\in\mathsf{F}})$, *and let* $p \in (0,1)$, $\varepsilon > 0$. *If* $w_{n,p}(\mathsf{B}) \geq C(r)\varepsilon^{-2}\log n$ *for a sufficiently large constant* $C(r) > 0$, *then*

$$\mu_Q\big(\{A \in \mathcal{A}_{n,r} : \|A - Q\|_\mathsf{B}^* \geq \varepsilon p\}\big) \leq \exp(-c\varepsilon^2 p^{d_\star+2} n^r).$$

Proof
Let $\mathbb{P}_Q$ be a probability measure under which $\boldsymbol{A}$ has distribution μ_Q, and let $\mathbb{E}_Q$ be the associated expectation. From Hoeffding's inequality we get that for any $T \in \mathcal{T}_\mathsf{B}$,

$$\mathbb{P}_Q\big(|\langle \boldsymbol{A} - Q, T\rangle| > \varepsilon p\|T\|_\mathsf{B}\big) \leq 2\exp\Big(-\frac{c\varepsilon^2 p^2\|T\|_\mathsf{B}^2}{\|T\|_2^2}\Big) \leq 2\exp\big(-c'\varepsilon^2 p^2\|T\|_\mathsf{B}\big).$$

Recalling (5.7), by the union bound we have that for any $L > 0$,

$$\begin{aligned}
&\mathbb{P}_Q\big(\exists T \in \mathcal{T}_\mathsf{B} : \|T\|_\mathsf{B} = L, |\langle \boldsymbol{A} - Q, T\rangle| > \varepsilon p\|T\|_\mathsf{B}\big)\\
&\quad\leq 2\exp\big(O_r(\log n)\max_{\mathsf{f}\in\mathsf{F}}\{n^{|\mathsf{f}|-r}p^{d_\mathsf{f}-d_\star}\}L - c'\varepsilon^2 p^2 L\big)\\
&\quad\leq 2\exp\big(-(c'/2)\varepsilon^2 p^2 L\big),
\end{aligned}$$

assuming that $C(r)$ in our assumption (6.11) is sufficiently large. Since the set of possible values for $\|T\|_\mathsf{B}$ is contained in a union of $|\mathsf{F}| + 1 \leq 2^r$ arithmetic progressions in $[n^r p^{d_\star}, n^r]$ of step at least 1, and $n^r p^{d_\star+2} \geq w_{n,p}(\mathsf{B})$, the claim follows from another union bound and summing the at most 2^r geometric series (and we can take $C(r)$ larger if necessary to absorb the prefactor 2^r). □

Proof of Lemma 6.2
From Lemma 6.3, identifying $\mathcal{A}_{n,r}$ with $\{0,1\}^{\binom{n}{r}}$, we get that for any $\mathcal{E} \subseteq \mathcal{A}_{n,r}$ and $Q \in \mathcal{Q}_{n,r}$,

$$\log \mu_p(\mathcal{E}) \geq -\mathrm{I}_p(Q) + \log \mu_Q(\mathcal{E}) - O\Big(\frac{n^{r/2}|\log(p \wedge (1-p))|}{\mu_Q(\mathcal{E})^{1/2}}\Big).$$

Taking $\mathcal{E} = \bigcap_{e\in\mathsf{E}} \mathcal{E}_e$ with $\mathcal{E}_e = \{A \in \mathcal{A}_{n,r} : \|A - Q\|^*_{\mathsf{B}(e)} < \varepsilon p\}$, from Lemma 6.4 we have $\mu_Q(\mathcal{E}_e) \geq 1 - 1/(2|\mathsf{E}|)$ if $C(r)$ is sufficiently large. From the union bound we get that $\mu_Q(\mathcal{E}) \geq 1/2$, and the claim follows upon substituting this estimate in the above display. □

7. Proof of Theorem 1.1—Upper tail

In this section we establish (1.8). We also show how (1.9) can be established along similar lines under some alternative assumptions on p and $\underline{H}$. The proof of (1.9) under the assumption $p \gg n^{-1/\Delta}$ is quite different and is given in Section 9.

To lighten notation, we will first present the proof of (1.8) for the case $m = 1$ in Section 7.1, and then describe in Section 7.2 the simple modifications that are needed for the general case.

7.1. Upper bound for the upper tail probability (case $m = 1$)

In this subsection we prove the following proposition, yielding (1.8) for the case $m = 1$.

PROPOSITION 7.1
For any r-graph H and $\delta, \xi > 0$, assuming that

$$np^{\Delta'(H)} \log(1/p) > C_2 \xi^{-3} \log n \tag{7.1}$$

for sufficiently large $C_2(H, \delta) > 0$, we have

$$\log \mathbb{P}\big(t_p(H, \boldsymbol{A}) > 1 + \delta\big) \leq -(1 - \xi)\Phi_{n,p}(H, \delta - \xi). \tag{7.2}$$

We need the following lemma, which one obtains by the same lines as in [61, Theorem 2.2] (for the lower bound they do not use the stated assumption that $p \gg n^{-1/\Delta(H)}$, and their upper-bound assumption on p is also not necessary, as one can use Lemma 9.5 below in place of their Lemma 4.7 to get sufficient control on I_p).

LEMMA 7.2
For any $n \in \mathbb{N}$, $p \in (0, 1)$, $u > 0$, and r-graph H with $\Delta(H) \geq 2$,

$$\Phi_{n,p}(H, u) \gtrsim_{H,u} n^r p^{\Delta(H)} \log(1/p).$$

Moreover, for $u \geq 1$,

$$\Phi_{n,p}(H, u) \gtrsim_H u^{\Delta(H)/\mathsf{e}(H)} n^r p^{\Delta(H)} \log(1/p).$$

In the remainder of this subsection, we set $\Delta := \Delta(H)$ and denote

$$R_{n,p} := n^r p^{\Delta} \log(1/p). \tag{7.3}$$

The following claim, giving the tail probability for the event under which we can apply Theorem 2.15, will be proved together with Proposition 7.1 by induction on the number of edges in H.

CLAIM 7.3
There exists $C_2'(H) > 0$ such that for any $L > C_2'(H)$, if

$$np^{\Delta'(H)} \log(1/p) \geq C_2' L^3 \log n \tag{7.4}$$

then if $\Delta \geq 2$,

$$\mathbb{P}\big(t_p(H, \boldsymbol{A}) \geq L\big) \leq \exp(-cL^{1/\mathsf{e}(H)} R_{n,p}) \tag{7.5}$$

and when $\Delta = 1$,

$$\mathbb{P}\big(t_p(H, \boldsymbol{A}) \geq L\big) \leq \exp(-cL^{1/\mathsf{e}(H)} n^r p \log L). \tag{7.6}$$

The key to the proof of Proposition 7.1 is that Theorem 2.15 only requires control over *proper* subgraphs of H as given by Claim 7.3, and thus can be guaranteed inductively. (Observe that the conclusion of Claim 7.3 can be essentially obtained from the conclusion of Proposition 7.1 and the crude upper bound in Lemma 7.2; however, the constant C_2' in Claim 7.3 is independent of $L = 1 + \delta$, which will be important for closing the inductive argument.)

Proof of Proposition 7.1
We proceed by induction on the number of edges in H. For the case $\mathsf{e}(H) = 1$, Claim 7.3 follows from a standard tail bound for the binomial distribution, and Proposition 7.1 follows immediately from Lemma 6.1 as the superlevel set in this case is a (convex) half-space. From the fact that $t_p(H_1 \cup H_2, Q) = t_p(H_1, Q) t_p(H_2, Q)$ for H a disjoint union of two graphs H_1, H_2 we further obtain the case $\Delta = 1$ for Proposition 7.1 and Claim 7.3.

Assume now that $\Delta \geq 2$ and that Proposition 7.1 and Claim 7.3 hold when $\mathsf{e}(H) \leq \ell - 1$ for some $\ell \geq 2$. Consider now an r-graph H with $\mathsf{e}(H) = \ell$.

Recalling (3.14), for each $e \in \mathsf{E}(H)$ we fix a dominating weighted base $\mathsf{B}(e)$ as in (2.15) with $d'_{\mathsf{F}(e)}(e) \leq \Delta'(H_k)$, and form the base system $\mathbb{B} = \{\mathsf{B}(e)\}_{e \in \mathsf{E}(H)}$. This implies

$$w_{n,p}(\mathbb{B}) \geq \min_{e \in \mathsf{E}(H)} \min_{\mathsf{f} \in \mathsf{F}(e)} (np^{\Delta'(H)})^{r - |\mathsf{f}|} \geq np^{\Delta'(H)} \tag{7.7}$$

since by definition we have $|\mathsf{f}| < r$ for any element f of any base (note that we may alternatively express $d'_{\mathsf{f}}(e)$ in (3.14) as $(d^H(e) - d^H_{\mathsf{f}}(e) + 2)(r - |\mathsf{f}|)$).

We denote sublevel sets by

$$\mathscr{L}_H(L) := \big\{Q \in \mathcal{Q}_{n,r} : t_p(H,Q) \le L\big\}, \quad L > 0, \tag{7.8}$$

and additionally denote

$$\mathscr{L}_{<H}(L) := \bigcap_{F \subsetneq H} \mathscr{L}_F(L). \tag{7.9}$$

With C_2, C_2' to be determined over the course of the proof, consider for now arbitrary $\delta, \xi > 0$, and additional parameters $L_0 > C_2'(H)$, $K \ge 1$, and $\varepsilon > 0$. We may assume that $\xi < \delta/2$. Taking $C_2'(H) \ge \max_{F \subsetneq H} C_2'(F)$, from the induction hypothesis and the union bound we have

$$\mathbb{P}\big(\boldsymbol{A} \notin \mathscr{L}_{<H}(L_0)\big) \lesssim_H \exp(-cL_0^{1/\mathsf{e}(H)} R_{n,p}). \tag{7.10}$$

(Here we used our assumption $\Delta \ge 2$: note that for $F \subset H$ of max-degree 1, while (7.6) loses a factor $\log(1/p)$ in the exponent as compared to (7.5), we gain a factor $p^{1-\Delta}$.) Now set

$$\mathcal{E} := \mathscr{A}_{n,r} \cap \mathscr{L}_H(1+\delta)^c \cap \mathscr{L}_{<H}(L_0).$$

From the previous bound and Theorem 3.1 we have

$$\begin{aligned}\mathbb{P}\big(\boldsymbol{A} \in \mathscr{L}_H(1+\delta)^c\big) &\lesssim_H \exp(-cL_0^{1/\mathsf{e}(H)} R_{n,p}) + \mathbb{P}(\boldsymbol{A} \in \mathcal{E}) \\ &\lesssim_H \exp(-cL_0^{1/\mathsf{e}(H)} R_{n,p}) + \exp(-c_0 K R_{n,p}) \\ &\quad + \exp\Big(\frac{C_1 K R_{n,p} \log n}{\varepsilon^2 n p^{\Delta'(H)} \log(1/p)} - \mathrm{I}_p\big((\mathcal{E})_{\mathbb{B},\varepsilon p}\big)\Big)\end{aligned}$$

(recall the shorthand notation (6.10)). From Theorem 2.15 it follows that

$$(\mathcal{E})_{\mathbb{B},\varepsilon p} \subseteq \mathscr{L}_H\big(1 + \delta - O_H(L_0\varepsilon)\big)^c,$$

and hence

$$\mathrm{I}_p\big((\mathcal{E})_{\mathbb{B},\varepsilon p}\big) \ge \Phi_{n,p}\big(H, \delta - O_H(\varepsilon L_0)\big). \tag{7.11}$$

Substituting this bound into the previous bound, we get that for any $K \ge 1$, $L_0 > C_2'(H)$, $\delta \ge 2\xi > 0$, and $\varepsilon < c(H)\xi/L_0$ for $c(H) > 0$ sufficiently small,

$$\begin{aligned}\mathbb{P}\big(\boldsymbol{A} \in \mathscr{L}_H(1+\delta)^c\big) &\lesssim_H \exp(-cL_0^{1/\mathsf{e}(H)} R_{n,p}) + \exp(-c_0 K R_{n,p}) \\ &\quad + \exp\Big(\frac{C_1 K R_{n,p} \log n}{\varepsilon^2 n p^{\Delta'(H)} \log(1/p)} - \Phi_{n,p}(H, \delta - \xi)\Big) \\ &=: \mathrm{(I)} + \mathrm{(II)} + \mathrm{(III)}.\end{aligned} \tag{7.12}$$

Now to establish (7.2) under the assumption (7.1), from Lemma 7.2 we can take L_0, K sufficiently large depending on H, δ to make the terms (I) and (II) in (7.12) negligible. Fixing such L_0, K, we can then fix $\varepsilon = c'\xi$ for sufficiently small $c'(H,\delta) > 0$, so that, together with (7.7) and our assumption (7.1), by taking $C_2(H,\delta)$ sufficiently large we make the first term in the exponential of (III) at most $\frac{\xi}{2}\Phi_{n,p}(H,\delta-\xi)$, and (7.2) follows.

For (7.5), in (7.12) we take $L_0 = 2L \geq C_2'(H)$, $\delta = L-1$, $\xi = 1/2$ (say), $K = 2L^{1/\mathsf{e}(H)}$, $\varepsilon = c''/10L$ for sufficiently small $c''(H) > 0$, and combining with (7.7) and (7.4), the claim then follows upon taking $C_2'(H)$ sufficiently large. □

7.2. Upper bound for the upper-tail probability (general case)

For the case of general $m \in \mathbb{N}$, we follow similar lines as in Section 7.1 with some minor modifications. The proof is now by induction on $\ell := \max_k \mathsf{e}(H_k)$. The case $\ell = 1$ is handled exactly as before (the half-spaces being intersected have parallel boundaries). For $\ell \geq 2$, we fix a dominating weighted base $\mathsf{B}(e)$ for each $k \in [m]$ and $e \in \mathsf{E}(H_k)$, and (7.7) now holds with the minimum now taken over all edges in all graphs H_k. In place of (7.3) we now take

$$R_{n,p} := n^r p^{\min_k \Delta(H_k)} \log(1/p). \tag{7.13}$$

The bounds in Lemma 7.2 extend to $\Phi_{n,p}(\underline{H},\underline{\delta})$ by restricting the infimum to a single superlevel set $\{t_p(H_k,\cdot) \geq 1+\delta_k\}$ for which $\Delta(H_k) = \min_\ell \Delta(H_\ell)$. We apply the upper-LDP bound of Theorem 3.1(a) with

$$\mathcal{E} := \mathcal{A}_{n,r} \cap \mathcal{L}_{\underline{H}}(\mathbf{1}+\underline{\delta})^c \cap \mathcal{L}_{<\underline{H}}(L_0),$$

where for $\underline{L} \in \mathbb{R}_+^m$,

$$\mathcal{L}_{\underline{H}}(\underline{L}) := \bigcap_{k\in[m]} \mathcal{L}_{H_k}(L_k), \qquad \mathcal{L}_{<\underline{H}}(L_0) := \bigcap_{k\in[m]} \bigcap_{F\subsetneq H_k} \mathcal{L}_F(L_0). \tag{7.14}$$

For this choice of $\mathcal{E}$, Theorem 2.15 implies $(\mathcal{E})_{\mathbb{B},\varepsilon p} \subseteq \mathcal{L}_{\underline{H}}(\mathbf{1}+\underline{\delta} - O_{\underline{H}}(L_0\varepsilon))^c$, and the rest of the argument proceeds as before. □

7.3. Lower bound for the upper-tail probability

In this subsection we show how the bound (1.9) easily follows from Theorems 3.1 and 2.15 under some alternative assumptions. As in the proof of (1.8) we present only the case $m = 1$ to lighten notation, but the argument extends to general m in a straightforward way, following similar modifications as in Section 7.2. The proof assuming $p \gg n^{-1/\Delta(H)}$ is quite different and is given in Section 9.

PROPOSITION 7.4
In the setting of Theorem 1.1, the bound (1.9) holds if $p \gg n^{-\min(\frac{1}{\Delta'(H)}, \frac{r}{2\Delta(H)})} \log n$ *and if we further assume that either (a)* $p \geq n^{-1/\Delta(H)}$ *or (b)* H *is regular.*

This is an immediate consequence of Lemma 7.2 and the following, together with the assumption $p \gg n^{-1/\Delta'(H)} \log n$ (via (3.17)).

LEMMA 7.5
Let H *be an* r*-graph, and let* $\delta > \xi > 0$*. Suppose that there exists an* H*-dominating base system* $\mathbb{B} = \{\mathsf{B}(e)\}_{e \in \mathsf{E}(H)}$ *with weights as in (2.15) such that* $\min_{e \in \mathsf{E}(H)} w_{n,p}(\mathsf{B}(e)) \geq C(H,\delta)\xi^{-2} \log n$ *for a sufficiently large constant* $C(H,\delta) > 0$*. Then if* $p \geq n^{-1/\Delta(H)}$*, or if* $p \geq n^{-r/\Delta(H)}$ *and* H *is regular, we have*

$$\log \mathbb{P}\big(t_p(H, \boldsymbol{A}) \geq 1 + \delta\big) \geq -\Phi_{n,p}(H, \delta + \xi) - O\big(1 + n^{r/2}\big|\log\big(p \wedge (1-p)\big)\big|\big).$$

(To deduce case (b) in Proposition 7.4 note that $n^{-1/\Delta'(H)} \geq n^{-r/\Delta(H)}$ from (1.7), and the assumption $p \gg n^{-r/(2\Delta(H))}$ implies that the error term in Lemma 7.5 is $o(R_{n,p}) = o(\Phi_{n,p}(H,\delta))$.)

Proof
Setting $\mathcal{E} := \mathcal{A}_{n,r} \cap \mathcal{L}_H(1+\delta)^c$ and letting $\varepsilon = c\xi$ for $c = c(H,\delta) > 0$ to be taken sufficiently small, from Theorem 3.1(b) we have

$$\mathbb{P}\big(t_p(H, \boldsymbol{A}) \geq 1 + \delta\big) \geq -\mathrm{I}_p\big((\mathcal{E})^\circ_{\mathbb{B},\varepsilon p}\big) - O\big(1 + n^{r/2}\big|\log\big(p \wedge (1-p)\big)\big|\big), \quad (7.15)$$

assuming that $C(H,\delta)$ is sufficiently large depending on c and the constant $C_0'(r)$ from Theorem 3.1(b). It only remains to show that

$$\mathrm{I}_p\big((\mathcal{E})^\circ_{\mathbb{B},\varepsilon p}\big) \leq \Phi_{n,p}(H, \delta + \xi) \quad (7.16)$$

if c is sufficiently small. Letting $1 \leq K = O_{H,\delta}(1)$ to be chosen later, we first argue that

$$(\mathcal{E})^\circ_{\mathbb{B},\varepsilon p} \supseteq \mathcal{L}_H(1+\delta+\xi)^c \cap \mathcal{L}_{<H}(K). \quad (7.17)$$

Indeed, letting Q_0 be an arbitrary element of the left-hand side, for any $A \in \mathcal{U}_{\mathbb{B}}(Q_0, \varepsilon p)$ we have from Theorem 2.15 applied to $\mathcal{C} = \mathcal{U}_{\mathbb{B}}(Q_0, \varepsilon p)$ that

$$t_p(H, A) \geq t_p(H, Q_0) - O_{H,\delta}(\varepsilon) \geq 1 + \delta + \xi - O_{H,\delta}(\varepsilon) \geq 1 + \delta$$

if c is sufficiently small, and (7.17) follows. Next, by considering $Q \in \mathcal{Q}_{n,r}$ of the form $Q(I) = 1$ whenever $I \cap [\lfloor Lp^{\Delta(H)} n \rfloor] \neq \emptyset$ and otherwise $Q(I) = p$, one easily checks that

$$\Phi_{n,p}(H,\gamma) \lesssim_{H,\gamma} n^r p^{\Delta(H)} \log(1/p) \tag{7.18}$$

for any fixed $\gamma > 0$, assuming $p \geq n^{-1/\Delta(H)}$ and taking $L = L(H,\gamma)$ sufficiently large. If H is regular, then we get the same bound for $p \geq n^{-r/\Delta(H)}$ by considering Q of the form $Q(I) = 1$ for $I \subset [\lfloor Lp^{\Delta(H)/r} n \rfloor]$ and otherwise $Q(I) = p$. From Lemma 7.2, we can hence fix K sufficiently large depending on H and δ so that

$$\mathrm{I}_p\big(\mathscr{L}_F(K)^c\big) = \Phi_{n,p}(F, K-1) > \Phi_{n,p}(H, 2\delta)$$

for every subgraph F of H. Since $\Phi_{n,p}(H,2\delta) \geq \Phi_{n,p}(H,\delta+\xi)$, it follows that

$$\Phi_{n,p}(H,\delta+\xi) = \mathrm{I}_p\big(\mathscr{L}_H(1+\delta+\xi)^c\big) = \mathrm{I}_p\big(\mathscr{L}_H(1+\delta+\xi)^c \cap \mathscr{L}_{<H}(K)\big)$$

and (7.16) follows upon combining the above with (7.17). □

8. Proof of Theorem 1.1—Lower tail

Recall our notation $\mathscr{L}_H(u) = \{Q \in \mathcal{Q}_{n,r} : t_p(H,Q) \leq u\}$ and $\mathscr{L}_{<H}(u) = \bigcap_{F \subset H} \mathscr{L}_F(u)$. We will make frequent use of the shorthand notation (6.10). We need the following elementary estimate for the lower-tail optimization problem (1.6) (for a proof, see [60, Lemma 22]).

LEMMA 8.1
For any r-graph H and $\delta \in (0,1)$, $\Psi_{n,p}(H,\delta) = \mathrm{I}_p(\mathscr{L}_H(1-\delta)) \gtrsim_{H,\delta} n^r p$.

Since $\Psi_{n,p}(\underline{H},\underline{\delta}) \geq \Psi_{n,p}(H_k,\delta_k)$ for any $k \in [m]$, it immediately follows that $\Psi_{n,p}(\underline{H},\underline{\delta}) \gtrsim_{\underline{H},\underline{\delta}} n^r p$ for any $\underline{H}$ and $\underline{\delta} \in (0,1)^m$.

8.1. Upper bound on the lower-tail probability

As in the proof of the upper bound for the upper tail, we will combine the upper LDP of Theorem 3.1 with the counting lemma (Theorem 2.15). However, whereas for the upper tail we needed to argue by induction on the number of edges in H in order to satisfy the crude bound assumption (2.20) for subgraphs, here we can reduce to this assumption with a simple application of the FKG inequality.

LEMMA 8.2
Recalling the notation (7.14), there exists $C = C(\underline{H}) > 0$ such that

$$\mathbb{P}\big(\boldsymbol{A} \in \mathscr{L}_{\underline{H}}(\mathbf{1}-\underline{\delta})\big) \leq 2\mathbb{P}\big(\boldsymbol{A} \in \mathscr{L}_{\underline{H}}(\mathbf{1}-\underline{\delta}) \cap \mathscr{L}_{<\underline{H}}(C)\big).$$

Proof
Since the sets $\mathscr{L}_{\underline{H}}(\mathbf{1}-\underline{\delta})\}$ and $\mathscr{L}_{<\underline{H}}(C)\}$ are monotone subsets of the hypercube, we have by the FKG inequality that

$$\begin{aligned}\mathbb{P}\big(\boldsymbol{A} \in \mathcal{L}_{\underline{H}}(\mathbf{1}-\underline{\delta})\big) &= \mathbb{P}\big(\boldsymbol{A} \in \mathcal{L}_{\underline{H}}(\mathbf{1}-\underline{\delta}) \cap \mathcal{L}_{<\underline{H}}(C)\big) + \mathbb{P}\big(\boldsymbol{A} \in \mathcal{L}_{\underline{H}}(\mathbf{1}-\underline{\delta}), \boldsymbol{A} \notin \mathcal{L}_{<\underline{H}}(C)\big) \\ &\le \mathbb{P}\big(\boldsymbol{A} \in \mathcal{L}_{\underline{H}}(\mathbf{1}-\underline{\delta}) \cap \mathcal{L}_{<\underline{H}}(C)\big) + \mathbb{P}\big(\boldsymbol{A} \in \mathcal{L}_{\underline{H}}(\underline{\delta})\big)\mathbb{P}\big(\boldsymbol{A} \notin \mathcal{L}_{<\underline{H}}(C)\big).\end{aligned}$$

Rearranging yields

$$\mathbb{P}\big(\boldsymbol{A} \in \mathcal{L}_H(1-\delta)\big) \le \frac{\mathbb{P}(\boldsymbol{A} \in \mathcal{L}_{\underline{H}}(\mathbf{1}-\underline{\delta}) \cap \mathcal{L}_{<\underline{H}}(C))}{1-\mathbb{P}(\boldsymbol{A} \notin \mathcal{L}_{<\underline{H}}(C))}.$$

Now for any $k \in [m]$ and subgraph $F \subset H_k$, from Markov's inequality we get that $\mathbb{P}(\boldsymbol{A} \notin \mathcal{L}_F(C)) \lesssim 1/C$. The claim follows by applying this together with the union bound over $F \subset H_k$, $k \in [m]$ and taking C sufficiently large depending on $\underline{H}$. □

We proceed with the proof of the upper bound in (1.10). Let $c = c(\underline{H}) > 0$ to be taken sufficiently small, and set $\varepsilon := c\xi$. By our lower-bound assumption on p and (3.17), we can select a collection $\mathbb{B} = \{\mathsf{B}(e)\}_{e \in \bigcup_{k=1}^m \mathsf{E}(H_k)}$ of dominating weighted bases over the edges of each H_k with weights as in (2.15), such that for any fixed $W_0 = W_0(\underline{H}, \delta, \xi)$ and all n sufficiently large,

$$w_{n,p}\big(\mathsf{B}(e)\big) \ge W_0 \varepsilon^{-2} \log n. \tag{8.1}$$

With $C = C(\underline{H})$ as in Lemma 8.2, we denote $\mathcal{E} := \mathcal{A}_{n,r} \cap \mathcal{L}_{\underline{H}}(\mathbf{1}-\underline{\delta}) \cap \mathcal{L}_{<\underline{H}}(C)$. Taking W_0 larger than the constant $C_0(r)$ from Theorem 2.13, we can apply Theorem 3.1(a) with $\Delta = 1$ and some $1 \le K = O_H(1)$ to be chosen later to bound

$$\begin{aligned}\mathbb{P}(\boldsymbol{A} \in \mathcal{E}) \le -\min\big\{&c_0 K n^r p \log(1/p) - O_{\mathbb{B}}(1), \\ &\mathrm{I}_p\big((\mathcal{E})_{\mathbb{B},\varepsilon p}\big) - O_{\mathbb{B}}\big(K W_0^{-1} n^r p \log(1/p)\big)\big\}.\end{aligned} \tag{8.2}$$

Now we claim that

$$(\mathcal{E})_{\mathbb{B},\varepsilon p} \subseteq \mathcal{L}_{\underline{H}}(\mathbf{1}-\underline{\delta}+\xi\,\mathbf{1}) \tag{8.3}$$

if c is sufficiently small. Indeed, letting $A_0 \in \mathcal{E}$ be arbitrary, applying Theorem 2.15 with $\mathcal{C} = \mathcal{U}_{\mathbb{B}}(Q, \varepsilon p)$, we have that for each $1 \le k \le m$,

$$t_p(H_k, Q) \le t_p(A_0) + O_{\underline{H}}(\varepsilon) \le 1 - \delta_k + \xi$$

if c is sufficiently small, and (8.3) follows. Thus, $\mathrm{I}_p((\mathcal{E})_{\mathbb{B},\varepsilon p}) \ge \Psi_{n,p}(\underline{H}, \underline{\delta} - \xi\,\mathbf{1})$. From Lemma 8.1 (and the remark that follows it) we can fix K sufficiently large and then fix W_0 sufficiently large such that the minimum in (8.2) is attained by the second argument, and such that the error term in the second argument is at most $\xi \Psi_{n,p}(\underline{H}, \underline{\delta} - \xi\,\mathbf{1})$. The lower bound on $\mathrm{LT}_{n,p}(\underline{H}, \underline{\delta})$ in (1.10) follows. □

8.2. Lower bound for the lower-tail probability

The upper bound on $\mathrm{LT}_{n,p}(H,\delta)$ in (1.10) is a consequence of Lemma 8.1 and the following, together with our assumption on p (via (3.17)).

PROPOSITION 8.3

Let H be an r-graph, and let $\delta \in (0,1)$, $\xi \in (0, 1-\delta)$. Assume that there exists a collection $\mathbb{B} = \{\mathsf{B}(e)\}_{e\in\mathsf{E}(H)}$ of weighted bases such that $\min_{e\in\mathsf{E}(H)} w_{n,p}(\mathsf{B}(e)) \geq C(r)\xi^{-2}\log n$ for a sufficiently large constant $C(r) > 0$. Then

$$\log \mathbb{P}\big(t_p(H,\boldsymbol{A}) \leq 1-\delta\big) \geq -\Psi_{n,p}(H,\delta+\xi) - O\big(1+n^{r/2}\big|\log\big(p\wedge(1-p)\big)\big|\big).$$

To establish the proposition we need two lemmas. In the following we denote by $\mathcal{Q}_{n,r}^{\leq p}$ the set of all $Q \in \mathcal{Q}_{n,r}$ with entries uniformly bounded by p.

LEMMA 8.4

For any sequence of r-graphs $\underline{H} = (H_1,\ldots,H_m)$ and $\underline{\delta} \in (0,1)^m$, we have $\mathrm{I}_p(\mathcal{L}_{\underline{H}}(\mathbf{1}-\underline{\delta})) = \mathrm{I}_p(\mathcal{L}_{\underline{H}}(\mathbf{1}-\underline{\delta}) \cap \mathcal{Q}_{n,r}^{\leq p})$.

Proof

Clearly the left-hand side is bounded by the right-hand side. For the reverse inequality, let $Q \in \mathcal{L}_{\underline{H}}(\mathbf{1}-\underline{\delta})$ be arbitrary. Since $t_p(H_k,\cdot)$ is monotone increasing in each coordinate we have $t_p(H_k, Q\wedge p) \leq t_p(H_k,Q)$ for each k, so $Q\wedge p \in \mathcal{L}_{\underline{H}}(\mathbf{1}-\underline{\delta})$. Furthermore, since I_p is monotone increasing under increasing $Q(I) \in [p,1]$ for any $I \in \binom{[n]}{r}$ with all other coordinates held fixed, we have that $\mathrm{I}_p(Q) \geq \mathrm{I}_p(Q\wedge p)$, and the claim follows. □

LEMMA 8.5

There exists $c = c(\underline{H}) > 0$ such that for any $\xi > 0$,

$$\mathcal{L}_{\underline{H}}(\mathbf{1}-\underline{\delta}-\xi\,\mathbf{1}) \cap \mathcal{Q}_{n,r}^{\leq p} \subseteq \big(\mathcal{A}\cap\mathcal{L}_{\underline{H}}(\mathbf{1}-\underline{\delta})\big)^{\circ}_{\mathbb{B},c\xi p}.$$

Proof

We write $\mathcal{E} := \mathcal{A}\cap\mathcal{L}_{\underline{H}}(\mathbf{1}-\underline{\delta})$. Let $c > 0$ to be taken sufficiently small depending on H, and put $\varepsilon = c\xi$. Fixing an arbitrary $Q \in \mathcal{L}_{\underline{H}}(\mathbf{1}-\underline{\delta}-\xi\,\mathbf{1}) \cap \mathcal{Q}_{n,r}^{\leq p}$, our aim is to show that $\mathcal{U}_{\mathbb{B}}(Q,\varepsilon p) \subseteq \mathcal{L}_{\underline{H}}(\mathbf{1}-\underline{\delta})$. Since $Q \in \mathcal{Q}_{n,r}^{\leq p}$ we have by monotonicity that $t_p(F,Q) \leq t_p(F,p) \leq 1$ for every r-graph F. In particular, $Q \in \mathcal{L}_{<\underline{H}}(1)$. Applying Theorem 2.15 with $\mathcal{C} = \mathcal{U}_{\mathbb{B}}(Q,\varepsilon p)$ and $Q_0 = Q$ we have

$$t_p(H_k,Q) - t_p(H_k,A) \lesssim_H \varepsilon$$

for each $k \in [m]$ and every $A \in \mathcal{U}_{\mathbb{B}}(Q, \varepsilon p)$. From the triangle inequality it follows that $t_p(H_k, A) \le 1 - \delta_k - \xi + O_{\underline{H}}(\varepsilon)$ for every such k and A, and the claim follows by taking $\varepsilon = c\xi$ for some $c = c(\xi) > 0$ sufficiently small. □

Proof of Proposition 8.3
Setting $\varepsilon = c\xi$ for $c(\underline{H}) > 0$ sufficiently small and denoting $\mathcal{E} := \mathcal{A} \cap \mathcal{L}_{\underline{H}}(\mathbf{1} - \underline{\delta})$, from Theorem 3.1(b) it suffices to show that

$$\mathrm{I}_p\big((\mathcal{E})^\circ_{\mathbb{B},\varepsilon p}\big) \le \Psi_{n,p}(\underline{H}, \underline{\delta} + \xi\,\mathbf{1}) = \mathrm{I}_p\big(\mathcal{L}_{\underline{H}}(\mathbf{1} - \underline{\delta} - \xi\,\mathbf{1})\big).$$

From Lemma 8.4 the right-hand side is equal to $\mathrm{I}_p(\mathcal{L}_{\underline{H}}(\mathbf{1} - \underline{\delta} - \xi\,\mathbf{1}) \cap \mathcal{Q}_{n,r}^{\le p})$, and from Lemma 8.5 this is bounded below by the left-hand side above as long as c is sufficiently small. □

9. Alternative lower-bound argument: Concentration under the tilted law

In this section we establish the bound (1.9) of Theorem 1.1 under the stated assumption $p \gg n^{-1/\Delta_{\min}}$. We have already shown in Section 7.3 how the lower bound can be established using the general LDP lower bound of Theorem 3.1(b) together with the counting lemma. Theorem 3.1(b) in turn is based on concentration of the B^*-norms under the tilted Erdős–Rényi laws μ_Q (see Lemma 6.4). Here we instead show that the homomorphism counts themselves are concentrated under the tilted law. In fact, we show this for the more general signed homomorphism counts of Theorem 2.15, which also includes *induced* homomorphism counts as a special case.

The bound (1.9) under the assumption $1 > p \gg n^{-1/\Delta(H)}$ is a consequence of the following. For the upper-bound assumption on p, recall Remark 1.2.

PROPOSITION 9.1
In the setting of Theorem 1.1, if $n^{-1/\Delta_{\max}} \ll p \le \min_k (1+\delta_k)^{-1/\mathsf{e}(H_k)}$, then for any fixed $\xi > 0$ and all n sufficiently large,

$$\mathbb{P}\big(t_p(H_k, \boldsymbol{A}) > 1 + \delta_k, 1 \le k \le m\big) \ge \frac{1}{2} \exp\big(-(1+\xi)\Phi_{n,p}(\underline{H}, \underline{\delta} + \xi\,\mathbf{1})\big).$$

The main step is to show that for any r-graph H and any $Q \in \mathcal{Q}_n$ satisfying

$$\mathrm{I}_p(Q) \lesssim n^r p^{\Delta(H)} \log(1/p),$$

we have that if $p \ge n^{-1/\Delta(H)}$, then the homomorphism counting functional concentrates under μ_Q, in the sense that

$$\mathrm{Var}_Q\big(t_p(H, \boldsymbol{A})\big) \ll \big(\mathbb{E}_Q t_p(H, \boldsymbol{A})\big)^2.$$

We prove the following more general statement for signed homomorphisms (recall the definitions from (4.1)).

PROPOSITION 9.2
Let $n^{-1/\Delta(H_+)} \le p \le 1-\varepsilon_0$ for some $\varepsilon_0 > 0$, and let $Q \in \mathcal{Q}_{n,r}$ be such that

$$\mathrm{I}_p(Q) \le K n^r p^{\Delta(H_+)} \log(1/p)$$

for some $K \in (0,\infty)$. Then

$$\mathrm{Var}_Q\big(\hom(\mathcal{H}, \boldsymbol{A})\big) \lesssim_{K,\mathcal{H},\varepsilon_0} \frac{\hom(\mathcal{H}, Q)^2}{n p^{\Delta(H_+)}}.$$

Taking Proposition 9.2 as given, we now establish Proposition 9.1 by a tilting argument.

Proof of Proposition 9.1
Fix $\xi > 0$. We may assume that $\xi \le 1$. The argument is slightly simpler when either p is fixed or all of the H_k have the same max-degree. We thus introduce a cutoff parameter $p_0(\xi, \underline{H}, \underline{\delta}) > 0$ to be taken sufficiently small, and let $U \subseteq [m]$ be defined to be $U := \{k \in [m] : \Delta(H_k) = \Delta_{\min}\}$ when $p < p_0$, and $U := [m]$ when $p \ge p_0$, where we write $\Delta_{\min} := \min_k \Delta(H_k)$. We write $\underline{H}_U$, $\underline{\delta}_U$ for the restriction of the sequences $\underline{H}$, $\underline{\delta}$ to the indices in U.

For the main step of the argument, we find an element $Q_\star \in \mathcal{Q}_{n,r}$ for which

$$\mathrm{I}_p(Q_\star) \le (1+\xi/2)\Phi_{n,p}(\underline{H}_U, \underline{\delta}_U + \xi\,\mathbf{1}) \tag{9.1}$$

and such that the joint upper-tail event is likely under the tilted law $\mu_{Q_\star}$, specifically:

$$\mathbb{P}_{Q_\star}\Big(\boldsymbol{A} \in \bigcap_{k=1}^{m} \mathcal{L}_{H_k}(1+\delta_k)^c\Big) \ge 3/4 \tag{9.2}$$

(recalling the notation (7.14)). We begin by fixing an arbitrary element $Q \in \mathcal{Q}_{n,r}$ such that

$$t_p(H_k, Q) \ge 1 + \delta_k + \xi \quad \forall k \in U$$

and

$$\mathrm{I}_p(Q) \le (1+\xi/4)\Phi_{n,p}(\underline{H}_U, \underline{\delta}_U + \xi\,\mathbf{1}). \tag{9.3}$$

By considering $Q_0 \in \mathcal{Q}_{n,r}$ of the form $Q_0(I) = p + (1-p)1_{I \cap [n_0] \ne \emptyset}$ for $n_0 = \lfloor C n p^{\Delta_{\min}} \rfloor$, one easily verifies that $t_p(H_k, Q_0) \ge 2 + \delta_k \ge 1 + \delta_k + \xi$ for all $k \in U$ if $C = C(\underline{H}, \underline{\delta}) > 0$ is sufficiently large, and thus we see that the right-hand side above is $O_{\underline{H},\underline{\delta}}(n^r p^{\Delta_{\min}} \log(1/p))$. Hence,

$$\mathrm{I}_p(Q) \lesssim_{\underline{H},\underline{\delta}} n^r p^{\Delta_{\min}} \log(1/p). \tag{9.4}$$

Note that for each $k \in U$,

$$\mathbb{E}_Q \hom(H_k, \boldsymbol{A}) \geq \hom(H_k, Q). \tag{9.5}$$

Combining this with Proposition 9.2 applied with $\mathscr{H} = H_k$ (all edges receiving the positive sign) and $\varepsilon_0 = \varepsilon_0(\underline{\delta}, \mathscr{H}) = 1 - \min_k (1+\delta_k)^{-1/\mathrm{e}(H_k)} > 0$, we have

$$\mathrm{Var}_Q\big(t_p(H_k, \boldsymbol{A})\big) \ll \big(\mathbb{E}_Q t_p(H_k, \boldsymbol{A})\big)^2$$

for each $k \in U$. In particular, from Chebyshev's inequality and the union bound we get

$$\mathbb{P}_Q\Big(\boldsymbol{A} \in \bigcap_{k \in U} \mathscr{L}_{H_k}(1+\delta_k)^c\Big) \geq 1 - o(1).$$

For the case that $U = [m]$, this yields (9.2) with $Q_\star = Q$.

For the case that $p < p_0$ and not all $\Delta(H_k)$ are equal, we modify Q slightly. First, setting $n_1 = \lfloor C' n p^{\Delta_{\min}+1} \rfloor$, let Q_0' have entries $Q_0'(I) = p$ for all $I \subset [n_1+1, n]$, and otherwise $Q_0'(I) = 1$. By the same computation as for Q_0 above, we have that $t_p(H_k, Q_0') \geq 1 + \delta_k + \xi$ for each $k \in [m] \setminus U$ if $C'(\underline{H}, \underline{\delta})$ is sufficiently large. Now let $Q_\star := Q \vee Q_0'$ be the entrywise maximum of Q and Q_0'. By monotonicity we have that (9.2) holds. Moreover, since $Q_\star$ only differs from Q on $O_{\underline{H},\underline{\delta}}(n^r p^{\Delta_{\min}+1})$ entries we have

$$\mathrm{I}_p(Q_\star) \leq \mathrm{I}_p(Q) + O_{\underline{H},\underline{\delta}}\big(n^r p^{\Delta_{\min}+1} \log(1/p)\big).$$

On the other hand, for any $k \in U$ we can lower bound $\Phi_{n,p}(\underline{H}_U, \underline{\delta}_U + \xi \mathbf{1}) \geq \Phi_{n,p}(H_k, \delta_k + \xi)$, and combining with Lemma 7.2, the above, and (9.3) we get

$$\begin{aligned}\mathrm{I}_p(Q_\star) &\leq \big(1 + \xi/4 + O_{\underline{H},\underline{\delta}}(p)\big)\Phi_{n,p}(\underline{H}_U, \underline{\delta}_U + \xi \mathbf{1}) \\ &\leq (1+\xi/2)\Phi_{n,p}(\underline{H}_U, \underline{\delta}_U + \xi \mathbf{1})\end{aligned}$$

taking p_0 sufficiently small depending on ξ, $\underline{H}$, and $\underline{\delta}$, giving (9.1) as desired.

Setting $\mathscr{E} := \bigcap_{k=1}^m \mathscr{L}_H(1+\delta_k)^c$, we next observe that

$$\mathbb{E}\mathbb{I}(\boldsymbol{A} \in \mathscr{E}) = \mathbb{E}_{Q_\star}\mathbb{I}(\boldsymbol{A} \in \mathscr{E}) \exp\big(-W(\boldsymbol{A})\big),$$

where

$$W(\boldsymbol{A}) = \sum_{I \in \binom{[n]}{r}} \Big\{ \boldsymbol{A}(I) \log \frac{Q(I)}{p} + \big(1 - \boldsymbol{A}(I)\big) \log \frac{1-Q(I)}{1-p} \Big\}.$$

By Lemma 7.2, the random variable $W(\boldsymbol{A})$ has expectation

$$\mathbb{E}_{Q_\star}\big[W(\boldsymbol{A})\big] = \mathrm{I}_p(Q_\star) \gtrsim_{\underline{H},\underline{\delta}} n^r p^{\Delta_{\min}} \log(1/p)$$

and variance

$$\begin{aligned}
\mathrm{Var}_{Q_\star}\big(W(\boldsymbol{A})\big) &= \sum_I \Big(\log\frac{p}{Q_\star(I)} - \log\frac{1-p}{1-Q_\star(I)}\Big)^2 Q_\star(I)\big(1-Q_\star(I)\big) \\
&\le 4\sum_I \big((\log p)^2 + \big(\log(1-p)\big)^2\big) Q_\star(I)\big(1-Q_\star(I)\big) \\
&\quad + 4\sum_I \big(\big(\log Q_\star(I)\big)^2 + \big(\log\big(1-Q_\star(I)\big)\big)^2\big) Q_\star(I)\big(1-Q_\star(I)\big) \\
&\lesssim_{\underline{H},\underline{\delta}} n^r (\log p)^2
\end{aligned}$$

since we assumed that $1 - p \gtrsim_{\underline{H},\underline{\delta}} 1$. Thus,

$$\mathrm{Var}_{Q_\star}\big(W(\boldsymbol{A})\big) \ll \big(\mathbb{E}_{Q_\star}\big[W(\boldsymbol{A})\big]\big)^2.$$

In particular, for sufficiently large n, we have a subset $\mathcal{E}'$ of $\mathcal{E}$ with $\mathbb{P}_{Q_\star}(\mathcal{L}') \ge \mathbb{P}_{Q_\star}(\mathcal{E}) - 1/4 \ge 1/2$ such that for all $A \in \mathcal{E}'$, $W(A) \le (1+\xi/10)\,\mathrm{I}_p(Q_\star)$. Therefore,

$$\mathbb{P}(\boldsymbol{A} \in \mathcal{E}) \ge \mathbb{E}_{Q_\star}\mathbb{I}(\boldsymbol{A} \in \mathcal{E}')\exp\big(-W(\boldsymbol{A})\big) \ge \frac{1}{2}\exp\big(-(1+\xi/10)\,\mathrm{I}_p(Q_\star)\big),$$

and hence

$$\begin{aligned}
\mathbb{P}(\boldsymbol{A} \in \mathcal{E}) &\ge \frac{1}{2}\exp\big(-(1+\xi/10)(1+\xi/2)\Phi_{n,p}(\underline{H}_U, \underline{\delta}_U + \xi\,\mathbf{1})\big) \\
&\ge \frac{1}{2}\exp\big(-(1+\xi)\Phi_{n,p}(\underline{H}_U, \underline{\delta}_U + \xi\,\mathbf{1})\big) \\
&\ge \frac{1}{2}\exp\big(-(1+\xi)\Phi_{n,p}(\underline{H}, \underline{\delta} + \xi\,\mathbf{1})\big),
\end{aligned}$$

as desired. □

Remark 9.3

As was pointed out to us by a referee, for $r \ge 3$ and for $r = 2$ under the stronger assumption that $(np^{\Delta_{\min}})^2 \gg \log(1/p)$, one can prove Proposition 9.1 using much weaker control of the event $\boldsymbol{A} \in \bigcap_{k=1}^m \mathcal{L}_{H_k}(1+\delta_k)^c$ (e.g., via Markov's inequality), together with stronger control of the random variable $W(\boldsymbol{A})$, which is a sum of independent random variables. We have decided to proceed via Proposition 9.2 as it may be of independent interest.

It remains to establish Proposition 9.2. We begin by collecting a few lemmas. The first is a Brascamp–Lieb-type generalization of Hölder's inequality that has been applied extensively in previous works analyzing the upper-tail optimization problem.

LEMMA 9.4 (Finner's inequality [41]; see also [65, Theorem 3.1])
For each $i \in [n]$, let Ω_i be a probability space with measure μ_i. Let $\mu = \bigotimes_{i=1}^n \mu_i$. Let $A_1, A_2, \dots, A_n$ be nonempty subsets of $[n] = \{1, 2, \dots, n\}$, and for $A \subseteq [n]$ let $\mu_A = \bigotimes_{i \in A} \mu_i$ and $\Omega_A = \prod_{i \in A} \Omega_i$. Let $f_i \in L^{q_i}(\Omega_{A_i}, \mu_{A_i})$ for each $i \le m$. Assume that $\sum_{i : j \in A_i} q_i^{-1} \le 1$ for all $j \le n$. Then we have

$$\int \prod_{i=1}^m f_i \, d\mu \le \prod_{i=1}^m \Big(\int |f_i|^{q_i} \, d\mu_{A_i} \Big)^{1/q_i}.$$

LEMMA 9.5
For any $p \in (0,1)$ and $0 \le x \le 1 - p$, we have $\mathrm{I}_p(p + x) \gtrsim x^2 \log(1/p)$.

Proof
The claim is trivial for $p \in [c, 1)$ for any fixed constant $c > 0$ by the uniform convexity of I_p, so we may assume that p is sufficiently small. Since $\mathrm{I}_p(p) = \mathrm{I}_p'(p) = 0$ and $\mathrm{I}_p''(x) \ge 1/x$, the claimed bound holds for $x \le 1/\log(1/p)$. For larger x one simply notes that $\mathrm{I}_p'(x) \gtrsim \log(1/p)$ for $x \ge \sqrt{p}$ (say). □

LEMMA 9.6
Let $\Delta \ge 2$, $K > 0$, and let G be an r-graph with maximum degree at most Δ. Suppose that $Q \in \mathcal{Q}_{n,r}$ has all entries in $[p, 1]$, and

$$\mathrm{I}_p(Q) \le K n^r p^\Delta \log(1/p).$$

Then

$$t_p(G, Q) \lesssim_{K,G} 1.$$

Proof
For economy of notation, we write $\boldsymbol{i} = (i_1, \dots, i_r)$. Let $Z(\boldsymbol{i}) = Q(\boldsymbol{i}) - p \in [0, 1-p]$. We have

$$\hom(G, Q) = \sum_{G' \subseteq G} p^{\mathsf{e}(G) - \mathsf{e}(G')} \hom(G', Z),$$

where G' ranges over subgraphs of G with the same vertex set.

Let $G' \subseteq G$. Then as the maximum degree of G (and hence G') is at most Δ, we have $\sum_{e \in \mathsf{E}(G') : v \in e} \frac{1}{\Delta} \le 1$. Thus, by Lemma 9.4,

$$\begin{aligned} n^{-\mathsf{v}(G')} \hom(G', Z) &= n^{-\mathsf{v}(G')} \sum_{\psi : \mathsf{V}(G) \to [n]} \prod_{e \in \mathsf{E}(G')} Z\big(\psi(e)\big) \\ &\le \prod_{e \in \mathsf{E}(G')} \Big(n^{-r} \sum_{\psi : e \to [n]} Z\big(\psi(e)\big)^{\Delta} \Big)^{1/\Delta} \\ &\le \Big(n^{-r} \sum_{\boldsymbol{i} \in [n]^r} Z(\boldsymbol{i})^2 \Big)^{\mathsf{e}(G')/\Delta}. \end{aligned}$$

By Lemma 9.5,

$$\sum_{\boldsymbol{i} \in [n]^r} Z(\boldsymbol{i})^2 \lesssim \frac{\mathrm{I}_p(Q)}{\log(1/p)}.$$

Thus,

$$\hom(G', Z) \le n^{\mathsf{v}(G')} O\Big(\frac{n^{-r} \mathrm{I}_p(Q)}{\log(1/p)} \Big)^{\mathsf{e}(G')/\Delta} \le O(K+1)^{\mathsf{e}(G')/\Delta} n^{\mathsf{v}(G)} p^{\mathsf{e}(G')},$$

and hence

$$\hom(G, Q) \le O(K+1)^{\mathsf{e}(G)/\Delta} n^{\mathsf{v}(G)} \sum_{G' \subseteq G} p^{\mathsf{e}(G) - \mathsf{e}(G') + \mathsf{e}(G')} \lesssim_{K,G} n^{\mathsf{v}(G)} p^{\mathsf{e}(G)},$$

as claimed. □

Below we will need to compare $\mathbb{E}_Q \hom(G, \boldsymbol{A})$ with $\hom(G, Q)$. By expanding the polynomial $\hom(G, \boldsymbol{A})$ and taking expectations, one sees that since the entries of Q lie in $[0,1]$, we have $\mathbb{E}_Q \hom(G, \boldsymbol{A}) \ge \hom(G, Q)$. The next lemma shows that a nearly matching upper bound holds. For a hypergraph G, let $\mathcal{S}(G)$ be the collection of hypergraphs G' such that there exists a surjective map Ψ from $\mathsf{V}(G)$ to $\mathsf{V}(G')$ such that $\mathsf{E}(G') = \{\Psi(e) : e \in \mathsf{E}(G)\}$ (without repeated elements). In particular, $G \in \mathcal{S}(G)$. We have

$$\mathbb{E}_Q \hom(G, \boldsymbol{A}) \le \sum_{G' \in \mathcal{S}(G)} \hom(G', Q). \tag{9.6}$$

The next lemma shows that the contribution from $G' \in \mathcal{S}(G)$ with $G' \ne G$ is negligible.

LEMMA 9.7

Let G be an r-graph with maximum degree at most Δ, and let $G' \in \mathcal{S}(G)$. Assume that $\mathsf{v}(G') < \mathsf{v}(G)$ and that $p \ge n^{-1/\Delta}$. If $Q \in \mathcal{Q}_{n,r}$ has all entries at least p and satisfies

$$\mathrm{I}_p(Q) \le K n^r p^\Delta \log(1/p),$$

then

$$\hom(G', Q) \lesssim_{K,G} \frac{n^{\mathsf{v}(G)} p^{\mathsf{e}(G)}}{np^\Delta}. \tag{9.7}$$

Furthermore, the same conclusion holds if all entries of Q are equal to some $q \le Kp$.

Proof
Since $G' \in \mathscr{S}(G)$, we have a surjection $\Psi : \mathsf{V}(G) \to \mathsf{V}(G')$ such that $\mathsf{E}(G') = \{\Psi(e) : e \in \mathsf{E}(G)\}$. Let $\Psi' : \mathsf{V}(G') \to \mathsf{V}(G)$ be any map such that $\Psi(\Psi'(v)) = v$ for all $v \in \mathsf{V}(G')$. Let $\overline{G}$ be the induced subgraph of G on $(\mathsf{V}(G'))$. Note that Ψ' gives a bijection between $\mathsf{V}(G')$ and $\mathsf{V}(\overline{G})$ such that $\Psi'^{-1}(\mathsf{E}(\overline{G})) \subseteq \mathsf{E}(G')$. For the first claim, observe that

$$\begin{aligned}
\hom(G', Q) &= \sum_{\psi:\mathsf{V}(G')\to[n]} \prod_{e\in\mathsf{E}(G')} Q\big(\psi(e)\big) \\
&\le \sum_{\psi:\mathsf{V}(G')\to[n]} \prod_{e\in\mathsf{E}(\overline{G})} Q\big(\psi\big(\Psi'^{-1}(e)\big)\big) \\
&= \sum_{\phi:\mathsf{V}(\overline{G})\to[n]} \prod_{e\in\mathsf{E}(\overline{G})} Q\big(\phi(e)\big) = \hom(\overline{G}, Q),
\end{aligned}$$

where in the second equality we let $\phi = \psi \circ \Psi'^{-1}$. Note that $\overline{G}$ has maximum degree at most Δ as it is an induced subgraph of G. By the above and Lemma 9.6, we have

$$\hom(G, Q) \le \hom(\overline{G}, Q) \lesssim_{K,\overline{G}} n^{\mathsf{v}(\overline{G})} p^{\mathsf{e}(\overline{G})}.$$

In the case that all entries of Q are equal to $q \le K'p$, we trivially have

$$\hom(G', Q) \lesssim_{G,K'} n^{\mathsf{v}(\overline{G})} p^{\mathsf{e}(\overline{G})}.$$

Thus, for both cases, to obtain (9.7) it suffices to show that

$$n^{\mathsf{v}(\overline{G})-\mathsf{v}(G)} p^{\mathsf{e}(\overline{G})-\mathsf{e}(G)} \le (np^\Delta)^{-1}.$$

By our assumptions that $p \ge n^{-1/\Delta}$ and $\mathsf{v}(G') = \mathsf{v}(\overline{G}) < \mathsf{v}(G)$, it suffices to show that

$$\Delta\big(\mathsf{v}(G) - \mathsf{v}(\overline{G})\big) \ge \mathsf{e}(G) - \mathsf{e}(\overline{G}).$$

Noting that

$$\Delta\big(\mathsf{v}(G) - \mathsf{v}(\overline{G})\big) = \Delta\big|\mathsf{V}(G) \setminus \Psi'\big(\mathsf{V}(G')\big)\big| \ge \big|\big\{e \in \mathsf{E}(G) : e \not\subset g\big(\mathsf{V}(G')\big)\big\}\big|$$

and

$$\mathsf{e}(G) - \mathsf{e}(\overline{G}) = \big|\big\{e \in \mathsf{E}(G) : e \not\subset \Psi'\big(\mathsf{V}(G')\big)\big\}\big|,$$

we obtain the desired conclusion. □

Proof of Proposition 9.2
From the Efron–Stein inequality we have that

$$\mathrm{Var}_Q\big[\hom(\mathcal{H}, \boldsymbol{A})\big] \le \frac{1}{2} \sum_{I \in \binom{[n]}{r}} \mathbb{E}_Q\big[\big(\hom(\mathcal{H}, \boldsymbol{A}) - \hom(\mathcal{H}, \boldsymbol{A}_I)\big)^2\big],$$

where $\boldsymbol{A}_I, \boldsymbol{A}_I^0, \boldsymbol{A}_I^1 \in \mathcal{A}_n$ are equal to $\boldsymbol{A}$ (up to the symmetry constraint) for $\boldsymbol{A}_I(I)$ being a Bernoulli$(Q(I))$ random variable independent of $\boldsymbol{A}$, whereas $\boldsymbol{A}_I^1(I) = 1$ and $\boldsymbol{A}_I^0(I) = 0$. Taking expectation over the variables $\boldsymbol{A}_I(I)$, we have the bound

$$\begin{aligned}\mathrm{Var}_Q\big[\hom(\mathcal{H}, \boldsymbol{A})\big] &\le \sum_I Q(I)\big(1 - Q(I)\big)\mathbb{E}_Q \Delta_I(\mathcal{H}, \boldsymbol{A})^2 \\ &\le \sum_I Q(I)\mathbb{E}_Q \Delta_I(\mathcal{H}, \boldsymbol{A})^2, \end{aligned} \tag{9.8}$$

where

$$\Delta_I(\mathcal{H}, \boldsymbol{A}) := \hom(\mathcal{H}, \boldsymbol{A}_I^1) - \hom(\mathcal{H}, \boldsymbol{A}_I^0).$$

Recalling the notation

$$A_\xi^e = \begin{cases} A & e \in \mathsf{E}(H_+), \\ 1 - A & e \in \mathsf{E}(H_-), \end{cases}$$

for $A \in \mathcal{A}_n$, we have

$$\begin{aligned}\Delta_I(\mathcal{H}, \boldsymbol{A}) &= \sum_{E_0 \subset \mathsf{E}(H)} \sum_{\substack{\phi : \mathsf{V}(H) \to [n] \\ \phi^{-1}(I) = E_0}} \prod_{e \in \mathsf{E}(H)} (\boldsymbol{A}_I^1)_\xi^e(\phi_e) - \prod_{e \in \mathsf{E}(H)} (\boldsymbol{A}_I^0)_\xi^e(\phi_e) \\ &= \sum_{E_0 \subset \mathsf{E}(H)} \sum_{\substack{\phi : \mathsf{V}(H) \to [n] \\ \phi^{-1}(I) = E_0}} [1_{E_0 \subseteq \mathsf{E}(H_+)} - 1_{E_0 \subseteq \mathsf{E}(H_-)}] \\ &\quad \times \prod_{e \in \mathsf{E}(H_+) \setminus E_0} \boldsymbol{A}(\phi_e) \prod_{e \in \mathsf{E}(H_-) \setminus E_0} \big(1 - \boldsymbol{A}(\phi_e)\big).\end{aligned}$$

Squaring, dropping the negative summands, and bounding terms $1 - \boldsymbol{A}(\phi_e)$ by 1, we obtain

$$\Delta_I(\mathcal{H}, \boldsymbol{A})^2 \leq \sum_{\substack{E_0, E_1 \subset \mathsf{E}(H_+) \\ \text{or } E_0, E_1 \subset \mathsf{E}(H_-)}} \sum_{\substack{\phi^0, \phi^1 : \mathsf{V}(H) \to [n] \\ \phi^a(e) = I \ \forall e \in E_a, a = 1,2}} \prod_{e \in \mathsf{E}(H_+) \setminus E_0} \boldsymbol{A}(\phi_e^0) \prod_{e \in \mathsf{E}(H_+) \setminus E_1} \boldsymbol{A}(\phi_e^1).$$

Combining the above with the FKG inequality, we get that (9.8) is bounded by

$$\mathbb{E}_Q \sum_I \boldsymbol{A}(I) \sum_{\substack{E_0, E_1 \subset \mathsf{E}(H_+) \\ \text{or } E_0, E_1 \subset \mathsf{E}(H_-)}} \sum_{\substack{\phi^0, \phi^1 : \mathsf{V}(H) \to [n] \\ \phi^a(e) = I \ \forall e \in E_a, a = 1,2}} \prod_{e \in \mathsf{E}(H_+) \setminus E_0} \boldsymbol{A}(\phi_e^0) \times \prod_{e \in \mathsf{E}(H_+) \setminus E_1} \boldsymbol{A}(\phi_e^1).$$

We can interpret the above sum as counting homomorphisms of graphs G_+ obtained from $\mathcal{H}$ as follows. Let H^0, H^1 be disjoint copies of H. For $a = 0, 1$, let $E_a \subset \mathsf{E}(H^a)$, and write V_a for the union of the edges in E_a; further, let H_+^a be a subgraph of H^a isomorphic to H_+ (with $\mathsf{V}(H_+^a) = \mathsf{V}(H^a)$ and edges corresponding to the positively labeled edges of H). Let $\widetilde{H}$ be the union of $H^0 \cup H^1$ with a disjoint edge $e_\star$, and define a surjection Ψ from $\mathsf{V}(\widetilde{H}) = \mathsf{V}(H^0) \cup \mathsf{V}(H^1) \cup e_\star$ to $V' := (\mathsf{V}(H^0) \setminus V_0) \cup (\mathsf{V}(H^1) \setminus V_1) \cup e_*$ as follows:

- On V' we take Ψ to be the identity map.
- On each V_a we take Ψ to be any surjection to $e_\star$ such that $\Psi(e) = e_\star$ for every $e \in E_a$.

Let G be the graph on V' that is the image of $\widetilde{H}$ under Ψ—that is, its edge set consists of the images $\Psi(e)$ of $e \in \mathsf{E}(\widetilde{H})$, removing any repetitions. We further let G_+ be the subgraph of G on the same vertex set V' with edges that are the images of the edges of $H_+^0 \cup H_+^1$, together with the edge $e_\star$. Letting $\mathcal{G}_\mathcal{H}$ be the collection of graphs G_+ over V' that can be obtained in this way, we have

$$\sum_I \boldsymbol{A}(I) \Delta_I(\mathcal{H}, \boldsymbol{A})^2 \leq \sum_{G_+ \in \mathcal{G}_\mathcal{H}} \hom(G_+, \boldsymbol{A}).$$

Combining the previous displays and applying (9.6), we have shown that

$$\begin{aligned} \operatorname{Var}_Q\big[\hom(\mathcal{H}, \boldsymbol{A})\big] &\leq \mathbb{E}_Q \sum_{G_+ \in \mathcal{G}_\mathcal{H}} \hom(G_+, \boldsymbol{A}) \\ &\leq \sum_{G_+ \in \mathcal{G}_\mathcal{H}} \sum_{F \in \mathcal{S}(G_+)} \hom(F, Q). \end{aligned} \tag{9.9}$$

Now consider an arbitrary $F \in \mathcal{S}(G_+)$ for some $G_+ \in \mathcal{G}_{\mathcal{H}}$. By definition, F is the image of G_+ under a surjection $\Psi' : V' \to \mathsf{V}(F)$ that maps edges to edges. Let F' be the subgraph over $\mathsf{V}(F)$ obtained by removing the edge $\Psi'(e_*)$, and observe that F' is the image under the restriction Ψ'' of $\Psi' \circ \Psi$ to $\mathsf{V}(H^0 \cup H^1)$ of a subgraph H' of $H^0_+ \cup H^1_+$ over the same vertex set. (Specifically, it is the image of the subgraph of $H^0_+ \cup H^0_-$ obtained by removing the edges $E_0 \cup E_1$.) Since Ψ'' is a surjection from $\mathsf{V}(H^0 \cup H^1)$ to $\mathsf{V}(F)$, we have that $F' \in \mathcal{S}(H^0_+ \cup H^0_-)$. Moreover,

$$\mathsf{v}(F') = \mathsf{v}(F) \le |V'| = 2\mathsf{v}(H) - |V_0| - |V_1| + r < 2\mathsf{v}(H) = \mathsf{v}(H^0_+ \cup H^1_+),$$

where the strict inequality follows from the fact that each of V_0 and V_1 contain at least one edge by assumption. Let Q' be defined by the coordinatewise maximum between Q and p. Then $\mathrm{I}_p(Q') \le \mathrm{I}_p(Q) \le Kn^r p^{\Delta(H_+)} \log(1/p)$ and

$$\hom(F, Q) \le \hom(F', Q) \le \hom(F', Q').$$

From Lemma 9.7 we have

$$\hom(F', Q') \lesssim_{K,\mathcal{H}} \frac{n^{\mathsf{v}(H^0_+ \cup H^1_+)} p^{\mathsf{e}(H^0_+ \cup H^1_+)}}{np^{\Delta(H_+)}} = \frac{n^{2\mathsf{v}(H)} p^{2\mathsf{e}(H_+)}}{np^{\Delta(H_+)}}.$$

Combining these bounds with (9.9) yields the claim. □

10. Other applications

10.1. Upper tails for induced homomorphism counts

For an r-graph H and $Q \in \mathcal{Q}_{n,r}$, the induced homomorphism count of H in Q is defined as

$$\mathrm{ind}(H, Q) = \sum_{\phi : \mathsf{V}(H) \to [n]} \prod_{e \in \mathsf{E}(H)} Q\big(\phi(e)\big) \prod_{e \in \binom{\mathsf{V}(H)}{r} \setminus \mathsf{E}(H)} \big(1 - Q\big(\phi(e)\big)\big).$$

As before, this definition extends to symmetric r-tensors.

For simplicity, we only consider the analogue of (1.8)–(1.9) for the case $m = 1$. Define

$$\mathrm{UT}^{\mathrm{ind}}_{n,p}(H, \delta) = -\log \mathbb{P}\big(\mathrm{ind}(H, \boldsymbol{G}) \ge (1+\delta) n^{\mathsf{v}(H)} p^{\mathsf{e}(H)} (1-p)^{\binom{\mathsf{v}(H)}{r} - \mathsf{e}(H)}\big)$$

and the corresponding upper-tail optimization problem

$$\Phi^{\mathrm{ind}}_{n,p}(H, \delta) = \inf\big\{\mathrm{I}_p(Q) : Q \in \mathcal{Q}_{n,r}, \\ \mathrm{ind}(H, Q) \ge (1+\delta) n^{\mathsf{v}(H)} p^{\mathsf{e}(H)} (1-p)^{\binom{\mathsf{v}(H)}{r} - \mathsf{e}(H)}\big\}.$$

THEOREM 10.1
The bounds (1.9) and (1.8) hold with $\mathrm{UT}^{\mathrm{ind}}_{n,p}(H,\delta)$, $\Phi^{\mathrm{ind}}_{n,p}(H,\delta)$ *in place of* $\mathrm{UT}_{n,p}(H,\delta)$, $\Phi_{n,p}(H,\delta)$ *(in the case* $m=1$*), under the same lower-bound assumptions on* p*, and also assuming that* $p \le p_0$ *for an arbitrary fixed* $p_0 \in (0,1)$.

For the lower bound on $\mathrm{UT}^{\mathrm{ind}}_{n,p}(H,\delta)$, we follow the proof of Proposition 7.1. The only difference is that in place of Theorem 2.15 we apply the generalized counting lemma Theorem 4.1 with the signed hypergraph $\mathcal{K} = (K^r_{\mathsf{v}(H)}, \xi)$, where $\xi(e) = +1$ if $e \in \mathsf{E}(H)$ and $\xi(e) = -1$ if $e \in \binom{\mathsf{V}(H)}{r} \setminus \mathsf{E}(H)$. The subgraphs $K_\pm$ induced by ξ are then defined by $\mathsf{V}(K_\pm) = V(H)$, $\mathsf{E}(K_+) = \mathsf{E}(H)$, and $\mathsf{E}(K_-) = \binom{\mathsf{V}(H)}{r} \setminus \mathsf{E}(H)$. We also use in the proof the fact that

$$\Phi^{\mathrm{ind}}_{n,p}(H,\delta) \gtrsim_{H,\delta} n^r p^{\Delta(H)} \log(1/p)),$$

which follows from the argument of [61, Theorem 2.2]. For $p = \omega(n^{-1/\Delta(H)})$, we obtain a matching upper bound

$$\Phi^{\mathrm{ind}}_{n,p}(H,\delta) \lesssim_{H,\delta} n^r p^{\Delta(H)} \log(1/p),$$

by fixing a subset J_0 of $[n]$ of size $\Theta_{H,\delta}(np^{\Delta(H)})$, and let Q be so that Q takes value $1/2$ on hyperedges which intersect J_0 and Q takes value p elsewhere. We can verify that $\mathrm{I}_p(Q) = \Theta_{H,\delta}(n^r p^{\Delta(H)} \log(1/p))$.

Next, we show the upper bound on $\mathrm{UT}^{\mathrm{ind}}_{n,p}(H,\delta)$, following the proof of Proposition 9.1. We highlight the main changes and additional steps. Let $Q \in \mathcal{Q}_{n,r}$ be such that

$$\hom(\mathcal{K}, Q) \ge (1+\delta+\xi) n^{\mathsf{v}(K)} p^{\mathsf{e}(K_+)} (1-p)^{\mathsf{e}(K_-)} \gtrsim n^{\mathsf{v}(K)} p^{\mathsf{e}(K_+)}$$

and

$$\mathrm{I}_p(Q) = \sum_{I \in \binom{[n]}{r}} \mathrm{I}_p\big(Q(I)\big) \le (1+\xi/4)\Phi^{\mathrm{ind}}_{n,p}(H,\delta+\xi) \lesssim n^r p^{\Delta(K_+)} \log(1/p).$$

In order to show that

$$\mathrm{Var}_Q\big(\hom(\mathcal{K}, \boldsymbol{A})\big) \ll \big(\mathbb{E}_Q \hom(\mathcal{K}, \boldsymbol{A})\big)^2, \tag{10.1}$$

by Proposition 9.2 it suffices to show that

$$\hom(\mathcal{K}, Q) \lesssim \mathbb{E}_Q \hom(\mathcal{K}, \boldsymbol{A}). \tag{10.2}$$

The right-hand side is bounded below by $\mathbb{E}_Q \mathrm{inj}(\mathcal{K}, \boldsymbol{A}) = \mathrm{inj}(\mathcal{K}, Q)$, where we write $\mathrm{inj}(\mathcal{K}, \cdot)$ for the count of injective signed homomorphisms. On the other hand, the

count of noninjective signed homomorphisms of $\mathcal{K}$ in Q is at most the count of noninjective homomorphisms of $K_+ = H$ in Q, for which Lemma 9.7 gives

$$\hom(H, Q) - \mathrm{inj}(H, Q) = o(n^{\mathsf{v}(H)} p^{\mathsf{e}(H)}).$$

Since $1 - p \geq 1 - (1+\delta)^{-1/\mathsf{e}(H)} \gtrsim_{\delta, H} 1$, the right-hand side is $o(\hom(\mathcal{K}, Q))$. Thus, $\hom(\mathcal{K}, Q) = (1 + o(1)) \,\mathrm{inj}(\mathcal{K}, Q)$, which establishes (10.2) and hence (10.1).

We can easily show that

$$W(\boldsymbol{A}) = \sum_I \Big\{ \boldsymbol{A}(I) \log \frac{Q(I)}{p} + \big(1 - \boldsymbol{A}(I)\big) \log \frac{1 - Q(I)}{1 - p} \Big\}$$

concentrates around its expectation under $\mathbb{P}_Q$. We then obtain the desired upper bound on $\mathrm{UT}^{\mathrm{ind}}_{n,p}(H, \delta)$ as in Proposition 9.1.

10.2. Lower tails for Sidorenko hypergraphs

We call H a *Sidorenko hypergraph* if

$$\frac{\hom(H, Q)}{n^{\mathsf{v}(H)}} \geq \hom(K^r_r, Q)^{\mathsf{e}(H)} \quad \forall Q \in \mathcal{Q}_{n,r},$$

where K^r_r is simply one hyperedge. For the case $r = 2$, a famous conjecture in extremal combinatorics by Erdős and Simonovits [75] and Sidorenko [73] states that all bipartite graphs are Sidorenko. This conjecture has been verified for a large family of bipartite graphs, including trees, even cycles, paths, hypercubes, and bipartite graphs with one vertex complete to the other side (see [24], [29], [76] and references therein). While a natural generalization of Sidorenko's conjecture to hypergraphs is false, it is known that many families of hypergraphs satisfy the Sidorenko property (see [76]).

Let H be a graph (so $r = 2$). Let $\hat{q}$ be so that $\hat{q}^{\mathsf{e}(H)} \leq (1 - \delta) p^{\mathsf{e}(H)}$, and let $q = \hat{q} \frac{n}{n-1}$. It is established in [31] that

$$\mathrm{LT}_{n,p}(H, \delta) \leq \big(1 + o(1)\big) \binom{n}{2} \mathrm{I}_p(q),$$

as long as $p = \omega(n^{-1/(2\Delta_2(H)-1)})$. Furthermore, if H is a Sidorenko graph, then the following nonasymptotic bound holds:

$$\mathrm{LT}_{n,p}(H, \delta) \geq \binom{n}{2} \mathrm{I}_p(q).$$

Our next theorem generalizes this result to r-uniform Sidorenko hypergraphs and improves on the range of p where the lower-tail asymptotics hold. We remark that

in [60] the reduction of the lower-tail asymptotics to the corresponding variational problem has been shown in an optimal range of sparsity for graphs ($r = 2$), where the reduction is also shown for hypergraphs satisfying appropriate degree conditions.

Denote by $\mathbb{E}_q$ the expectation with respect to the random r-graph where each hyperedge is independently included with probability q.

THEOREM 10.2
Let H be an r-graph. Assume that $p = \omega(n^{-1/\Delta(H)})$. Fix $\delta \in (0,1)$, let $\hat{q} = (1-\delta)^{1/\mathsf{e}(H)} p$, and let $q = \hat{q}\frac{n^r}{n\cdots(n-r+1)}$. Then

$$\mathrm{LT}_{n,p}(H,\delta) \le \big(1+o(1)\big)\binom{n}{r}\mathrm{I}_p\big(\big(1-o(1)\big)q\big). \tag{10.3}$$

Furthermore, if H is a Sidorenko hypergraph, then we have

$$\mathrm{LT}_{n,p}(H,\delta) \ge \binom{n}{r}\mathrm{I}_p(q). \tag{10.4}$$

Our result thus yields the lower-tail asymptotics as long as H is Sidorenko and $p = \omega(n^{-1/\Delta(H)})$. We remark that in the regime $p = \omega(n^{-1/\Delta(H)})$, we can verify that $\hat{q} = \Theta(p)$ so $q = \Theta(p)$. In the case $r = 2$, this improves the threshold in [31].

We turn to the proof of Theorem 10.2. We first give the proof of (10.3) following the proof of Proposition 9.1. Let $\xi > 0$ be any sufficiently small real number. We choose $\tilde{q} = q(1-\xi)$. We write $\mathbb{E}_{\tilde{q}}$ and $\mathrm{Var}_{\tilde{q}}$ for expectation and variance under the distribution of a random tensor $\boldsymbol{A}$ whose entries are independent and identically distributed Bernoulli($\tilde{q}$) variables. We first establish an analogue of Proposition 9.2 showing the concentration of $\hom(H,\boldsymbol{A})$ where $\boldsymbol{A}$ has independent Bernoulli($\tilde{q}$) entries. In particular, we show that

$$\mathrm{Var}_{\tilde{q}}\big(\hom(H,\boldsymbol{A})\big) \ll \big(\mathbb{E}_{\tilde{q}}\hom(H,\boldsymbol{A})\big)^2. \tag{10.5}$$

First, notice that $cp \le \tilde{q} \le p$ for some constant $c \in (0,1)$ depending only on δ. We have that

$$\mathbb{E}_{\tilde{q}}\hom(H,\boldsymbol{A}) \ge \big(1+o(1)\big)n^{\mathsf{v}(H)}\tilde{q}^{\mathsf{e}(H)}. \tag{10.6}$$

Indeed, by summing over the injective homomorphisms, we obtain

$$\mathbb{E}_{\tilde{q}}\hom(H,\boldsymbol{A}) \ge \big(1-o(1)\big)n^{\mathsf{v}(H)}\tilde{q}^{\mathsf{e}(H)}.$$

Recall that we denote by $J_{n,r} \in \mathscr{A}_{n,r}$ the tensor with all "off-diagonal" elements equal to 1. Following identically the proof of Proposition 9.2, we obtain that

$$\mathrm{Var}_{\tilde{q}}\big(\hom(H,\mathbf{A})\big) \lesssim_H \sum_{G\in\mathcal{S}(\tilde{H}):\mathsf{v}(G)<2\mathsf{v}(H)} \hom(G,\tilde{q}J_{n,r}),$$

where $\tilde{H}$ is the hypergraph obtained from two disjoint copies of H, and recall that $\mathcal{S}(\tilde{H})$ is the collection of hypergraphs G such that there exists a surjective map f from $\mathsf{V}(\tilde{H})$ to $\mathsf{V}(G)$ such that $\mathsf{E}(G) = f(\mathsf{E}(\tilde{H}))$. By Lemma 9.7 applied with $Q = \tilde{q}J_{n,r}$, we obtain that for each $G \in \mathcal{S}(\tilde{H})$,

$$\hom(G,\tilde{q}J_{n,r}) = o(n^{2\mathsf{v}(H)}p^{2\mathsf{e}(H)}).$$

Thus,

$$\mathrm{Var}_{\tilde{q}}\big(\hom(H,\mathbf{A})\big) = o(n^{2\mathsf{v}(H)}p^{2\mathsf{e}(H)}) = o\big(\big(\mathbb{E}_{\tilde{q}}\hom(H,\boldsymbol{A})\big)^2\big),$$

using (10.6), yielding (10.5). The concentration of

$$W(\boldsymbol{A}) = \sum_{\boldsymbol{i}}\Big\{\boldsymbol{A}(\boldsymbol{i})\log\frac{\tilde{q}}{p} + \big(1-\boldsymbol{A}(\boldsymbol{i})\big)\log\frac{1-\tilde{q}}{1-p}\Big\}$$

easily follows noting that

$$\mathbb{E}_{\tilde{q}}W(\boldsymbol{A}) \geq cn^r\Big(\tilde{q}\log\frac{\tilde{q}}{p} + (1-\tilde{q})\log\frac{1-\tilde{q}}{1-p}\Big)$$

and

$$\mathrm{Var}_{\tilde{q}}\big(W(\boldsymbol{A})\big) \leq Cn^r\tilde{q}(1-\tilde{q})\Big(\log\frac{p}{\tilde{q}} - \log\frac{1-p}{1-\tilde{q}}\Big)^2,$$

so as $\tilde{q} \in [cp,p]$, we have $\mathrm{Var}_{\tilde{q}}(W(\boldsymbol{A})) = o((\mathbb{E}_{\tilde{q}}W(\boldsymbol{A}))^2)$. Combining this with (10.5),we obtain (10.3) as in the proof of Proposition 9.1.

To establish (10.4) under the additional assumption that H is Sidorenko, we note that if $\hom(H,\boldsymbol{A}) \leq (1-\delta)p^{\mathsf{e}(H)}n^{\mathsf{v}(H)}$, then by the Sidorenko property, $\hom(K_r^r,\boldsymbol{A}) \leq \hat{q}$. Noting that

$$\hom(K_r^r,\boldsymbol{A}) = n^{-r}\sum_{\boldsymbol{i}\in[n]^r}\boldsymbol{A}(\boldsymbol{i}) = \frac{n\cdots(n-r+1)}{n^r}\binom{n}{r}^{-1}\sum_{\boldsymbol{i}}\boldsymbol{A}(\boldsymbol{i}),$$

(10.4) follows from basic properties of the binomial distribution.

Acknowledgments. We thank Bhaswar Bhattacharya, Ronen Eldan, Jacob Fox, Shirshendu Ganguly, Eyal Lubetzky, Wojciech Samotij and Yufei Zhao for helpful comments on an earlier version of the paper, as well as Yang P. Liu and Yufei Zhao for

their permission to use Figure 1. We additionally thank the anonymous referees for their feedback and helpful suggestions to improve the exposition.

Cook's work was partially supported by National Science Foundation (NSF) grant DMS-2154029. Dembo's work was partially supported by NSF grant DMS-1954337. Pham's work was partially supported by a Two Sigma Fellowship, a Clay Research Fellowship and a Stanford Science Fellowship.

References

[1] P. ALLEN, J. BÖTTCHER, J. SKOKAN, and M. STEIN, *Regularity inheritance in pseudorandom graphs*, Random Structures Algorithms **56** (2020), no. 2, 306–338. MR 4060348. DOI 10.1002/rsa.20851. *(889)*

[2] F. AUGERI, *Nonlinear large deviation bounds with applications to Wigner matrices and sparse Erdős-Rényi graphs*, Ann. Probab. **48** (2020), no. 5, 2404–2448. MR 4152647. DOI 10.1214/20-AOP1427. *(882)*

[3] T. AUSTIN, *The structure of low-complexity Gibbs measures on product spaces*, Ann. Probab. **47** (2019), no. 6, 4002–4023. MR 4038047. DOI 10.1214/19-aop1352. *(882)*

[4] J. BALOGH, R. MORRIS, and W. SAMOTIJ, *Independent sets in hypergraphs*, J. Amer. Math. Soc. **28** (2015), no. 3, 669–709. MR 3327533. DOI 10.1090/S0894-0347-2014-00816-X. *(879, 893)*

[5] A. BASAK, *Upper tail of the spectral radius of sparse Erdős-Rényi graphs*, Probab. Theory Related Fields **187** (2023), no. 3–4, 885–947. MR 4664587. DOI 10.1007/s00440-023-01232-6. *(882)*

[6] A. BASAK and R. BASU, *Upper tail large deviations of regular subgraph counts in Erdős-Rényi graphs in the full localized regime*, Comm. Pure Appl. Math. **76** (2023), no. 1, 3–72. MR 4544794. DOI 10.1002/cpa.22036. *(882, 883, 884)*

[7] B. B. BHATTACHARYA, S. BHATTACHARYA, and S. GANGULY, *Spectral edge in sparse random graphs: Upper and lower tail large deviations*, Ann. Probab. **49** (2021), no. 4, 1847–1885. MR 4260469. DOI 10.1214/20-aop1495. *(882)*

[8] B. B. BHATTACHARYA and S. GANGULY, *Upper tails for edge eigenvalues of random graphs*, SIAM J. Discrete Math. **34** (2020), no. 2, 1069–1083. MR 4083586. DOI 10.1137/18M1230852. *(882)*

[9] B. B. BHATTACHARYA, S. GANGULY, E. LUBETZKY, and Y. ZHAO, *Upper tails and independence polynomials in random graphs*, Adv. Math. **319** (2017), 313–347. MR 3695877. DOI 10.1016/j.aim.2017.08.003. *(876, 883, 887, 889, 901, 902)*

[10] S. BHATTACHARYA and A. DEMBO, *Upper tail for homomorphism counts in constrained sparse random graphs*, Random Structures Algorithms **59** (2021), no. 3, 315–338. MR 4295566. DOI 10.1002/rsa.21011. *(876, 877, 882, 883, 890)*

[11] C. BORGS, J. T. CHAYES, H. COHN, and L. LOVÁSZ, "Identifiability for graphexes and the weak kernel metric" in *Building Bridges, II: Mathematics of László Lovász*, Bolyai Soc. Math. Stud. **28**, Springer, Berlin, 2019, 29–157. MR 4297779. *(878)*

[12] C. BORGS, J. T. CHAYES, J. GAUDIO, S. PETTI, and S. SEN, *A large deviation principle for block models*, to appear in Combin. Probab. Comput., preprint, arXiv:2007.14508v1 [math.PR]. *(882)*

[13] C. BORGS, J. T. CHAYES, L. LOVÁSZ, V. T. SÓS, B. SZEGEDY, and K. VESZTERGOMBI, "Graph limits and parameter testing" in *STOC'06: Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, ACM, New York, 2006, 261–270. MR 2277152. DOI 10.1145/1132516.1132556. *(878)*

[14] C. BORGS, J. T. CHAYES, L. LOVÁSZ, V. T. SÓS, and K. VESZTERGOMBI, *Convergent sequences of dense graphs, I: Subgraph frequencies, metric properties and testing*, Adv. Math. **219** (2008), no. 6, 1801–1851. MR 2455626. DOI 10.1016/j.aim.2008.07.008. *(878)*

[15] ———, *Convergent sequences of dense graphs, II: Multiway cuts and statistical physics*, Ann. of Math. (2) **176** (2012), no. 1, 151–219. MR 2925382. DOI 10.4007/annals.2012.176.1.2. *(878)*

[16] S. CHATTERJEE, *The missing log in large deviations for triangle counts*, Random Structures Algorithms **40** (2012), no. 4, 437–451. MR 2925306. DOI 10.1002/rsa.20381. *(881)*

[17] ———, *An introduction to large deviations for random graphs*, Bull. Amer. Math. Soc. (N.S.) **53** (2016), no. 4, 617–642. MR 3544262. DOI 10.1090/bull/1539. *(878)*

[18] ———, *Large Deviations for Random Graphs*, Lecture Notes in Math. **2197**, Springer, Cham, 2017. MR 3700183. DOI 10.1007/978-3-319-65816-2. *(878)*

[19] S. CHATTERJEE and A. DEMBO, *Nonlinear large deviations*, Adv. Math. **299** (2016), 396–450. MR 3519474. DOI 10.1016/j.aim.2016.05.017. *(876, 881, 887)*

[20] S. CHATTERJEE and P. DIACONIS, *Estimating and understanding exponential random graph models*, Ann. Statist. **41** (2013), no. 5, 2428–2461. MR 3127871. DOI 10.1214/13-AOS1155. *(876)*

[21] S. CHATTERJEE and S. R. S. VARADHAN, *The large deviation principle for the Erdős-Rényi random graph*, European J. Combin. **32** (2011), no. 7, 1000–1017. MR 2825532. DOI 10.1016/j.ejc.2011.03.014. *(876, 877, 878, 881, 883, 887)*

[22] A. COHEN ANTONIR, *The upper tail problem for induced 4-cycles in sparse random graphs*, Random Structures Algorithms **64** (2024), no. 2, 401–459. MR 4704274. DOI 10.1002/rsa.21187. *(882, 885)*

[23] A. COJA-OGHLAN, C. COOPER, and A. FRIEZE, *An efficient sparse regularity concept*, SIAM J. Discrete Math. **23** (2009/10), no. 4, 2000–2034. MR 2594968. DOI 10.1137/080730160. *(889, 893)*

[24] D. CONLON, J. FOX, and B. SUDAKOV, *An approximate version of Sidorenko's conjecture*, Geom. Funct. Anal. **20** (2010), no. 6, 1354–1366. MR 2738996. DOI 10.1007/s00039-010-0097-0. *(938)*

[25] D. CONLON, J. FOX, and Y. ZHAO, *Extremal results in sparse pseudorandom graphs*, Adv. Math. **256** (2014), 206–290. MR 3177293. DOI 10.1016/j.aim.2013.12.004. *(888, 889)*

[26] ———, *A relative Szemerédi theorem*, Geom. Funct. Anal. **25** (2015), no. 3, 733–762. MR 3361771. DOI 10.1007/s00039-015-0324-9. *(889)*

[27] D. CONLON and W. T. GOWERS, *Combinatorial theorems in sparse random sets*, Ann. of Math. (2) **184** (2016), no. 2, 367–454. MR 3548529. DOI 10.4007/annals.2016.184.2.2. *(888)*

[28] D. CONLON, W. T. GOWERS, W. SAMOTIJ, and M. SCHACHT, *On the KŁR conjecture in random graphs*, Israel J. Math. **203** (2014), no. 1, 535–580. MR 3273450. DOI 10.1007/s11856-014-1120-1. *(889, 893)*

[29] D. CONLON, J. H. KIM, C. LEE, and J. LEE, *Some advances on Sidorenko's conjecture*, J. Lond. Math. Soc. (2) **98** (2018), no. 3, 593–608. MR 3893193. DOI 10.1112/jlms.12142. *(938)*

[30] D. CONLON and J. LEE, *Finite reflection groups and graph norms*, Adv. Math. **315** (2017), 130–165. MR 3667583. DOI 10.1016/j.aim.2017.05.009. *(895)*

[31] N. COOK and A. DEMBO, *Large deviations of subgraph counts for sparse Erdős-Rényi graphs*, Adv. Math. **373** (2020), no. 107289. MR 4130460. DOI 10.1016/j.aim.2020.107289. *(877, 882, 883, 918, 938, 939)*

[32] ———, *Typical structure of sparse exponential random graph models*, to appear in Ann. Appl. Probab., preprint, arXiv:2208.06397v1 [math.PR]. *(876, 883)*

[33] B. DEMARCO and J. KAHN, *Tight upper tail bounds for cliques*, Random Structures Algorithms **41** (2012), no. 4, 469–487. MR 2993131. DOI 10.1002/rsa.20440. *(881)*

[34] ———, *Upper tails for triangles*, Random Structures Algorithms **40** (2012), no. 4, 452–459. MR 2925307. DOI 10.1002/rsa.20382. *(881)*

[35] A. DEMBO and E. LUBETZKY, *A large deviation principle for the Erdős-Rényi uniform random graph*, Electron. Commun. Probab. **23** (2018), no. 13. MR 3873786. DOI 10.1214/18-ECP181. *(882)*

[36] A. DEMBO and O. ZEITOUNI, *Large Deviations Techniques and Applications*, reprint of the 2nd ed., Stoch. Model. Appl. Probab. **38**, Springer, Berlin, 2010. MR 2571413. DOI 10.1007/978-3-642-03311-7. *(914)*

[37] S. DHARA and S. SEN, *Large deviation for uniform graphs with given degrees*, Ann. Appl. Probab. **32** (2022), no. 3, 2327–2353. MR 4430015. DOI 10.1214/21-aap1745. *(882)*

[38] R. ELDAN, *Gaussian-width gradient complexity, reverse log-Sobolev inequalities and nonlinear large deviations*, Geom. Funct. Anal. **28** (2018), no. 6, 1548–1596. MR 3881829. DOI 10.1007/s00039-018-0461-z. *(882, 883, 900)*

[39] R. ELDAN and R. GROSS, *Decomposition of mean-field Gibbs distributions into product measures*, Electron. J. Probab. **23** (2018), no. 35. MR 3798245. DOI 10.1214/18-EJP159. *(882)*

[40] ———, *Exponential random graphs behave like mixtures of stochastic block models*, Ann. Appl. Probab. **28** (2018), no. 6, 3698–3735. MR 3861824. DOI 10.1214/18-AAP1402. *(876)*

[41] H. FINNER, *A generalization of Hölder's inequality and some probability inequalities*, Ann. Probab. **20** (1992), no. 4, 1893–1901. MR 1188047. *(931)*

[42] A. FRIEZE and R. KANNAN, "The regularity lemma and approximation schemes for dense problems" in *Foundations of Computer Science (Burlington, 1996)*, IEEE

Comput. Soc. Press, Los Alamitos, 1996, 12–20. MR 1450598. DOI 10.1109/SFCS.1996.548459. *(877, 878, 889)*

[43] S. GERKE, M. MARCINISZYN, and A. STEGER, *A probabilistic counting lemma for complete graphs*, Random Structures Algorithms **31** (2007), no. 4, 517–534. MR 2362641. DOI 10.1002/rsa.20198. *(889, 893)*

[44] W. T. GOWERS, *Quasirandomness, counting and regularity for 3-uniform hypergraphs*, Combin. Probab. Comput. **15** (2006), no. 1–2, 143–184. MR 2195580. DOI 10.1017/S0963548305007236. *(895)*

[45] ———, *Hypergraph regularity and the multidimensional Szemerédi theorem*, Ann. of Math. (2) **166** (2007), no. 3, 897–946. MR 2373376. DOI 10.4007/annals.2007.166.897. *(895)*

[46] ———, *Decompositions, approximate structure, transference, and the Hahn-Banach theorem*, Bull. Lond. Math. Soc. **42** (2010), no. 4, 573–606. MR 2669681. DOI 10.1112/blms/bdq018. *(892)*

[47] J. GREBÍK and O. PIKHURKO, *Large deviation principles for block and step graphon random graph models*, preprint, arXiv:2101.07025v3 [math.PR]. *(882)*

[48] B. GREEN and T. TAO, *The primes contain arbitrarily long arithmetic progressions*, Ann. of Math. (2) **167** (2008), no. 2, 481–547. MR 2415379. DOI 10.4007/annals.2008.167.481. *(888)*

[49] B. GUNBY, *Upper tails of subgraph counts in sparse regular graphs*, preprint, arXiv:2010.00658. MR 4314974. *(877, 882, 890)*

[50] M. HAREL, F. MOUSSET, and W. SAMOTIJ, *Upper tails via high moments and entropic stability*, Duke Math. J. **171** (2022), no. 10, 2089–2192. MR 4484206. DOI 10.1215/00127094-2021-0067. *(882, 883, 884, 885)*

[51] S. JANSON, *Poisson approximation for large deviations*, Random Structures Algorithms **1** (1990), no. 2, 221–229. MR 1138428. DOI 10.1002/rsa.3240010209. *(881)*

[52] S. JANSON, K. OLESZKIEWICZ, and A. RUCIŃSKI, *Upper tails for subgraph counts in random graphs*, Israel J. Math. **142** (2004), no. 1, 61–92. MR 2085711. DOI 10.1007/BF02771528. *(881)*

[53] S. JANSON and L. WARNKE, *The lower tail: Poisson approximation revisited*, Random Structures Algorithms **48** (2016), no. 2, 219–246. MR 3449596. DOI 10.1002/rsa.20590. *(881)*

[54] R. KENYON, C. RADIN, K. REN, and L. SADUN, *Multipodal structure and phase transitions in large constrained graphs*, J. Stat. Phys. **168** (2017), no. 2, 233–258. MR 3667360. DOI 10.1007/s10955-017-1804-0. *(883)*

[55] ———, *The phases of large networks with edge and triangle constraints*, J. Phys. A **50** (2017), no. 43, art. ID 435001. MR 3718855. DOI 10.1088/1751-8121/aa8ce1. *(883)*

[56] ———, *Bipodal structure in oversaturated random graphs*, Int. Math. Res. Not. IMRN **2018**, no. 4, 1009–1044. MR 3801454. DOI 10.1093/imrn/rnw261. *(883)*

[57] J. H. KIM and V. H. VU, *Divide and conquer martingales and the number of triangles in a random graph*, Random Structures Algorithms **24**, no. 2, 166–174, 2004. MR 2035874. DOI 10.1002/rsa.10113. *(881)*

[58] Y. KOHAYAKAWA, "Szemerédi's regularity lemma for sparse graphs" in *Foundations of Computational Mathematics (Rio de Janeiro, 1997)*, Springer, Berlin, 1997, 216–230. MR 1661982. *(880)*

[59] Y. KOHAYAKAWA, T. ŁUCZAK, and V. RÖDL, *On K^4-free subgraphs of random graphs*, Combinatorica **17** (1997), no. 2, 173–213. MR 1479298. DOI 10.1007/BF01200906. *(893)*

[60] G. KOZMA and W. SAMOTIJ, *Lower tails via relative entropy*, Ann. Probab. **51** (2023), no. 2, 665–698. MR 4546629. DOI 10.1214/22-aop1610. *(883, 885, 924, 939)*

[61] Y. P. LIU and Y. ZHAO, *On the upper tail problem for random hypergraphs*, Random Structures Algorithms **58** (2021), no. 2, 179–220. MR 4201796. DOI 10.1002/rsa.20975. *(882, 883, 890, 900, 901, 919, 937)*

[62] L. LOVÁSZ, *Large Networks and Graph Limits*, Amer. Math. Soc. Colloq. Publ. **60**, Amer. Math. Soc., Providence, 2012. MR 3012035. DOI 10.1090/coll/060. *(878)*

[63] L. LOVÁSZ and B. SZEGEDY, *Limits of dense graph sequences*, J. Combin. Theory Ser. B **96** (2006), no. 6, 933–957. MR 2274085. DOI 10.1016/j.jctb.2006.05.002. *(878)*

[64] E. LUBETZKY and Y. ZHAO, *On replica symmetry of large deviations in random graphs*, Random Structures Algorithms **47** (2015), no. 1, 109–146. MR 3366814. DOI 10.1002/rsa.20536. *(882, 883, 893)*

[65] ———, *On the variational problem for upper tails in sparse random graphs*, Random Structures Algorithms **50** (2017), no. 3, 420–436. MR 3632418. DOI 10.1002/rsa.20658. *(883, 889, 901, 931)*

[66] T. ŁUCZAK, *On triangle-free random graphs*, Random Structures Algorithms **16** (2000), no. 3, 260–276. MR 1749289. DOI 10.1002/(SICI)1098-2418(200005)16:3<260::AID-RSA3>3.3.CO;2-Q. *(879)*

[67] M. MARKERING, *The large deviation principle for inhomogeneous Erdős-Rényi random graphs*, J. Theoret. Probab. **36** (2023), no. 2, 711–727. MR 4591855. DOI 10.1007/s10959-022-01181-1. *(882)*

[68] S. MUKHERJEE and B. B. BHATTACHARYA, *Replica symmetry in upper tails of mean-field hypergraphs*, Adv. in Appl. Math. **119** (2020), no. 102047. MR 4092987. DOI 10.1016/j.aam.2020.102047. *(883)*

[69] C. RADIN, K. REN, and L. SADUN, *A symmetry breaking transition in the edge/triangle network model*, Ann. Inst. Henri Poincaré D **5** (2018), no. 2, 251–286. MR 3813216. DOI 10.4171/AIHPD/54. *(883)*

[70] A. RAZ, *Upper tail bounds for cycles*, SIAM J. Discrete Math. **34** (2020), no. 3, 1559–1586. MR 4124632. DOI 10.1137/19M1258827. *(881)*

[71] D. SAXTON and A. THOMASON, *Hypergraph containers*, Invent. Math. **201** (2015), no. 3, 925–992. MR 3385638. DOI 10.1007/s00222-014-0562-8. *(879, 893)*

[72] M. SCHACHT, *Extremal results for random discrete structures*, Ann. of Math. (2) **184** (2016), no. 2, 333–365. MR 3548528. DOI 10.4007/annals.2016.184.2.1. *(888)*

[73] A. SIDORENKO, *A correlation inequality for bipartite graphs*, Graphs Combin. **9** (1993), no. 2, 201–204. MR 1225933. DOI 10.1007/BF02988307. *(938)*

[74] M. ŠILEIKIS and L. WARNKE, *Upper tail bounds for stars*, Electron. J. Combin. **27** (2020), no. 1, art. ID 1.67. MR 4093168. DOI 10.37236/8493. *(881, 902)*

[75] M. SIMONOVITS, "Extremal graph problems, degenerate extremal problems, and supersaturated graphs" in *Progress in Graph Theory (Waterloo, 1982)*, Academic Press, Toronto, 1984, 419–437. MR 0776819. *(938)*

[76] B. SZEGEDY, *An information theoretic approach to Sidorenko's conjecture*, preprint, arXiv:1406.6738v3 [math.CO]. *(938)*

[77] E. SZEMERÉDI, "On sets of integers containing no k elements in arithmetic progression" in *Proceedings of the International Congress of Mathematicians (Vancouver, 1974), Vol. 2*, Canad. Math. Congr., Montreal, 1975, 503–505. MR 0422191. *(888)*

[78] P. TURÁN, *On an extremal problem in graph theory*, Mat. Fiz. Lapok **48** (1941), 436–452. MR 0018405. *(888)*

[79] S. R. S. VARADHAN, *Asymptotic probabilities and differential equations*, Comm. Pure Appl. Math. **19** (1966), 261–286. MR 0203230. DOI 10.1002/cpa.3160190303. *(878)*

[80] J. YAN, *Nonlinear large deviations: Beyond the hypercube*, Ann. Appl. Probab. **30** (2020), no. 2, 812–846. MR 4108123. DOI 10.1214/19-AAP1516. *(882)*

[81] Y. ZHAO, *On the lower tail variational problem for random graphs*, Combin. Probab. Comput. **26** (2017), no. 2, 301–320. MR 3603970. DOI 10.1017/S0963548316000262. *(883)*

Cook

Department of Mathematics, Duke University, Durham, North Carolina, USA;
nickcook@math.duke.edu

Dembo

Department of Mathematics, Stanford University, Stanford, California, USA;
adembo@stanford.edu

Pham

Department of Mathematics, Stanford University, Stanford, California, USA;
huypham@stanford.edu