Structure of Lower Tails in Sparse Random Graphs

Random Structures & Algorithms

WILEY

RESEARCH ARTICLE OPEN ACCESS

Structure of Lower Tails in Sparse Random Graphs

Byron Chin

Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

Correspondence: Byron Chin (byronc@mit.edu)

Received: 11 February 2024 | **Revised:** 30 June 2025 | **Accepted:** 25 July 2025

Funding: This work was supported by the National Science Foundation (GRFP 2141064).

Keywords: large deviations | lower tails | random graphs | typical structure

ABSTRACT

We study the typical structure of a sparse Erdős–Rényi random graph conditioned on the lower tail subgraph count event. We show that in certain regimes, a typical graph sampled from the conditional distribution resembles the entropy minimizer of the mean field approximation in the sense of both subgraph counts and cut norm. The main ingredients are an adaptation of an entropy increment scheme of Kozma and Samotij, and a new stability for the solution of the associated entropy variational problem. The proof can be interpreted as a structural application of the new probabilistic hypergraph container lemma for sparser than average sets, and suggests a more general framework for establishing such typical behavior statements.

1 | Introduction

We study the Erdős–Rényi random graph conditioned on the large deviation event of a lower tail subgraph count. For a fixed graph H, let $N_H(G)$ be the number of copies of H (as a subgraph) in a graph G. We write $N_H(p)$ to denote the random variable $N_H(G(n,p))$, where $G(n,p)$ is a graph drawn from the Erdős–Rényi distribution. The key question is: Given that the lower tail event $\mathcal{L}_p(H,\eta) := \{G : N_H(G) \le \eta \mathbb{E}[N_H(p)]\}$ holds for some $\eta < 1$, what does a typical sample from the conditional distribution look like?

The study of this question was initiated by the seminal work of Chatterjee and Varadhan (2011), who proved a general large deviations principle for the Erdős–Rényi random graph with constant density, that is, for fixed $p \in (0,1)$. Their results imply the following two estimates, written below, which we first describe informally. The first statement (1.1) says that the leading order in the exponent of the lower tail probability is determined by the way to induce few copies of H with minimum entropic cost. The second statement (1.2) gives a more detailed description of the structure of the event, and says that the probability that a random sample with few copies of H looks very different from these entropy minimizers is exponentially small. They showed

$$-\log \mathbb{P}\left(\mathcal{L}_p(H,\eta)\right) = (1+o(1))\Phi_p(H,\eta), \tag{1.1}$$

$$\mathbb{P}\left(\delta_\square(G,M) > \varepsilon \mid \mathcal{L}_p(H,\eta)\right) \leq e^{-cn^2} \tag{1.2}$$

where $\Phi_p(H,\eta)$ is the value of a variational problem that minimizes the entropic cost over the lower tail event (see Definition 8 for the precise formulation), M is the set of minimizers to the variational problem $\Phi_p(H,\eta)$ in the space of graphons, and $\delta_\square$ is the cut distance. See (Chatterjee and Varadhan 2011) for a formal discussion of these expressions, or Section 2 for the definitions and their role in this paper. Much finer questions have also been studied in the case of dense graphs, determining the typical structure of large graphs with various fixed subgraph densities (i.e., Neeman et al. 2023; Kenyon et al. 2018; Radin 2018 and references within).

There are two barriers that prevent the extension of (Chatterjee and Varadhan 2011) to the case of sparse random graphs. First, their proof relies on Szemerédi's regularity lemma, which is ineffective at handling graphs with vanishing edge density. Second, and perhaps more significantly, the large deviation principle is formulated in terms of graphons, a natural limit object for dense graphs. The space of graphons is a compact topological space which nicely encodes similarity and distance between graphs, see (Lovász 2012) for a comprehensive survey. Theories for the limit of sequences of sparse graphs exist, for example, (Borgs et al. 2019), but there is no convenient limit object that captures all the same properties as a graphon. Consequently, it is not clear how a corresponding large deviation principle should be formulated. Despite these difficulties with generalizing the entire large deviations principle, it is possible to study these tail probabilities in a more specific manner.

For Equation (1.1), a natural first step is to replace Szemerédi's regularity lemma with the weak regularity lemma of Frieze and Kannan (1999) which can handle slowly vanishing densities $p \geq (\log n)^{-c}$. A breakthrough technique for computing large deviation probabilities for nonlinear functions of independent Bernoulli random variables was introduced by Chatterjee and Dembo (2016), which extends the range of densities to $p \geq n^{-\alpha(H)}$ for some explicit function $\alpha(H)$. This inspired a line of work developing the technique for arbitrary graphs and even uniform hypergraphs (Eldan 2018; Cook and Dembo 2020; Cook et al. 2021). Recently, a work of Kozma and Samotij (2023) showed (1.1) for $p \gg n^{-1/m_2(H)}$, the full range of densities for which this statement can be expected to hold, using an entropy increment technique. Here $m_2(H)$ refers to the 2-density of a graph H defined in Section 2. For sparser graphs, the lower tail probability is determined by a Poisson-type behavior, and Janson's inequality gives a tight bound. Janson and Warnke (2016) showed that this is true in a more general setting—see the references within for an overview of work in this even sparser regime. These lines of work present a fairly complete picture of the lower tail probability.

Extending the structural statement (1.2) to the sparse setting has seen far less progress. One reason for this is that the variational problem for lower tail subgraph counts has proven to be difficult to analyze. Very little is known about the set of minimizers, so sparse analogs of (1.2) are more difficult to show explicitly. Zhao (2017) showed that for η sufficiently close to 1 the constant function is the unique minimizer of $\Phi_p(H,\eta)$, whereas for η sufficiently close to 0, the constant function is no longer a minimizer. He studied a sparse limit of the variational problem, leading his results to hold for $p \to 0$. This remains the best-known range of parameters for which a solution to the lower tail variational problem is known. To the best of our knowledge, no analogs of (1.2) have been shown for the lower tail for any $p = o(1)$.

1.1 | Main Results

We provide a framework for deducing sparse analogs to (1.2)—the typical structure of a conditioned random graph—from a solution to the variational problem. We apply this argument in the setting of lower tail subgraph counts of $G(n,p)$, where all the ingredients are available to prove the complete result. We show that, conditioned on a random graph having few copies of one graph H', subgraph counts of another graph H are accurately predicted by the solution to the variational problem with high probability.

We work in a regime known as the replica symmetric regime, where the solution to the variational problem is known. We use $\eta_{H'}$ to refer to the constant from (Zhao 2017), above which the solution to the variational problem is known to be constant. Numerically, $\eta_{H'}$ is defined as the solution to a fixed-point equation and can be approximated. For example, η_{K_3} is 0.1012… —see (Zhao 2017, Section 5) for more estimates.

Theorem 1.1. *Fix any graphs H' and H. Let $G_{\mathcal{L}} \sim G(n,p)$ conditioned on $\mathcal{L}_p(H',\eta)$, where $\eta > \eta_{H'}$. Let $q = \eta^{1/e(H')}p$ where $e(H')$ is the number of edges of H' and $m = \max\{m_2(H), m_2(H')\}$. For any $\varepsilon > 0$ there exists an $L(\varepsilon,\eta,H,H')$ such that if $p \geq Ln^{-1/m}$,*

$$\mathbb{P}_{\mathcal{L}}(|N_H(G_{\mathcal{L}}) - \mathbb{E}[N_H(q)]| > \varepsilon\mathbb{E}[N_H(q)]) < \varepsilon.$$

Remark 1. For $p \gg n^{-1/m}$ carefully tracking asymptotics through the argument of (Kozma and Samotij 2023) should yield a decaying bound. However, as written, the proof only obtains a bound for an arbitrary fixed ε.

Here, $\mathbb{P}_{\mathcal{L}}$ refers to probability with respect to the conditional measure (see also Definition 3). As a consequence, we deduce the typical structure of the graph in the sense of the cut norm as well. The cut norm is a natural notion of distance between graphs that enjoys a set of nice properties when extended to the space of graphons (see Definition 4).

Corollary 1.2. *Fix any graph H. Let $G_{\mathcal{L}} \sim G(n,p)$ conditioned on $\mathcal{L}_p(H,\eta)$ where $\eta > \eta_H$. Let $q = \eta^{1/e(H)}p$. For any $\varepsilon > 0$ there exists an $L(\varepsilon,\eta,H)$ such that if $p \geq Ln^{-1/m_2(H)}$,*

$$\mathbb{P}_{\mathcal{L}}(\|G_{\mathcal{L}} - q\|_{\square} > \varepsilon p) < \varepsilon.$$

Our results extend (1.2) to the full range of densities for which the variational problem is relevant, and moreover establish typical structure in the stronger sense of subgraph counts. Indeed, for $p \ll n^{-1/m_2(H)}$ the large deviation rate is no longer controlled by the variational problem (see, e.g., Kozma and Samotij 2023, Theorem 2). In the sparse setting, it is known that subgraph counts control the cut norm, but the converse fails. Given our results about the typical structure of $G(n,p)$ conditioned on a lower tail event, one might wonder if an even stronger structural relationship holds. One possible formulation is the following:

Question 1. Fix any graph H, $\eta > \eta_H$ and let $q = \eta^{1/e(H)}p$. Let $p \gg n^{-1/m_2(H)}$ and $G_{\mathcal{L}} \sim G(n,p)$ conditioned on $\mathcal{L}_p(H,\eta)$. For any $\varepsilon > 0$, does there exist a coupling such that with high probability

$$G(n,(1-\varepsilon)q) \subset G_{\mathcal{L}} \subset G(n,(1+\varepsilon)q)?$$

A first step to investigate the above question is to study various statistics of the lower tail event that would follow from the existence of such a coupling.

Question 2. Let $G_{\mathcal{L}} \sim G(n,p)$ conditioned on $\mathcal{L}_p(H,\eta)$. What is the size of the largest clique in $G_{\mathcal{L}}$?

1.2 | Proof Overview

We combine a probabilistic version of the hypergraph container lemma with a new stability result to deduce the typical structure. This is motivated by recent progress combining the hypergraph container lemma with stability statements to obtain typical structure results for independent sets (see, e.g., Balogh et al. 2018).

Kozma and Samotij (2023) introduced a new entropy-based approach to estimating lower tail probabilities. They interpret their work as a "weak, probabilistic version of the hypergraph container lemma for sparser-than-average sets." The proof relies on an entropy increment to obtain approximate independence within a random graph conditioned on having few copies of H'. In particular, they iteratively decompose the lower tail measure as follows. Informally, at each step, either there are few dependencies among edges in copies of H' or there exists a small subset of edges that captures a non-trivial amount of entropy. Since the total entropy of the measure can be bounded, this can only be repeated a small number of times before few dependencies remain, leading to a small subset of edges that captures almost all of the dependence within copies of H'. This decomposes the lower tail measure into a mixture of measures that behave nicely with respect to copies of H'.

We leverage this iteration scheme in a novel way to decompose the measure with respect to a *different* graph H. Namely, in the lower tail measure for H', we find at each step either a few dependencies among edges in copies of H or a small subset of edges that captures a non-trivial amount of entropy. This decomposes the lower tail measure for H' into measures that behave nicely with respect to copies of any other graph H. Iterating this decomposition, we obtain approximate independence with respect to multiple subgraphs simultaneously. This multiple independence will be crucial to estimating subgraph counts later on.

We then extend the argument of Zhao (2017) for showing that the constant function is the unique entropy minimizer when the lower tail parameter η is sufficiently close to 1. He first shows that the optimizer must satisfy a point-wise Lagrange multiplier condition. Using this condition, he deduces that no entry of the optimizer can be too small. Finally, via a convexity argument, he concludes that the optimizer must be constant. We run a robust variant of the argument, allowing a bit of slack at each of these steps. Specifically, we consider the effect of increasing the optimizer in a fixed direction to show that the number of small entries must be small. Then, a convexity argument implies that nearly all the entries must be close to one another. This gives a corresponding stability result stating that near minimizers must be close to the constant function.

With these two ingredients in hand, our proof then proceeds as follows. Given the distribution of $G(n, p)$ conditioned on $\mathcal{L}_p(H', \eta)$, condition on a small set of edges to obtain approximate independence for copies of both H and H'. Then, with high probability, an independent sample from the marginal probabilities of each edge from this conditioned distribution will also satisfy a lower tail event for H'. This is because the independent marginals and the joint conditional measure behave similarly with respect to copies of H'. We then invoke our new stability result to deduce that the marginal edge probabilities must be close to the minimizer of the variational problem. From this, we deduce that the expected number of copies of H in this independent sample is close to the quantity predicted by the minimizer. Recalling our control on copies of H as well, we deduce that the expected number of copies of H under the joint distribution of edges is close to the independent sample from the marginals, and thus is also accurately predicted by the solution to the variational problem. Finally, to boost this estimate on the expectation to obtain concentration, we leverage the fact that the variance can be expressed in terms of subgraph counts and apply a second moment argument.

The bulk of this argument is, in fact, very general and easily allows for control over multiple subgraph counts as well as conditioning on multiple lower tail events. Furthermore, the main ideas can be phrased in the language of counting edges in hypergraphs, a flexible viewpoint exemplified in the transference principles of Conlon and Gowers (2016) and Schacht (2016) and the hypergraph container theorems (Balogh et al. 2015; Saxton and Thomason 2015). This would allow one to deduce a "typicality" statement for edge counts of two hypergraphs living on the same vertex set, provided we have the necessary stability of the variational problem. The primary bottleneck to these types of results appears to be the analysis of the entropy variational problem.

1.3 | Organization

The next section sets up the definitions and notation for the remaining arguments. In Section 3, we recall the main results of the entropy increment scheme to obtain approximate independence. In Section 4, we prove the new stability of the entropy variational problem. Section 5 combines the ingredients to complete the proof of the main theorem. We conclude in Section 6 by discussing our argument in a more general setup.

2 | Preliminaries

In this section, we introduce the primary definitions and notation that will be used throughout the paper. Some notation will be imported from (Kozma and Samotij 2023) and (Zhao 2017), and we will give a reference when this is the case. From here until Section 6, we will restrict our attention to the setting of subgraphs of random graphs. The more general setup will be introduced in the discussion in Section 6.

We let $G = (V, E)$ denote a graph and define the following statistics of fixed graphs.

Definition 1 (Graphs). Let $v(G) = |V|$ be the number of vertices and $e(G) = |E|$ be the number of edges of G. We use $N_H(G)$ to denote the number of copies of H as a subgraph of G. A key statistic is the 2-density of G, which is defined as the following

$$m_2(G) := \max\left\{\frac{e(F) - 1}{v(F) - 2} : F \subset G, e(F) \geq 2\right\}.$$

In this paper, we are interested in the behavior of random graphs, which will be notated according to the definition below.

Definition 2 (Random graphs). For a vector $\mathbf{q} \in [0, 1]^{\binom{n}{2}}$ let $G(n, \mathbf{q})$ be the random graph with vertex set $[n]$ where each edge $e = (i, j)$ is included independently with probability $q_{i,j} = q_e$. When $\mathbf{q} = p$ is a constant vector, we simply write $G(n, p)$, the well-known Erdős–Rényi random graph. We use $N_H(G)$ to represent the number of copies of H as a subgraph of G. We use $N_H(\mathbf{q})$ to represent $N_H(G(n, \mathbf{q}))$, the (random) number of copies of H in a sample from $G(n, \mathbf{q})$.

We are primarily interested in the lower tail subgraph count event, when the random graph has much fewer copies of a subgraph H than expected.

Definition 3 (Lower tail event). Let $\mathcal{L}_p(H,\eta)$ denote the event that $N_H(G) \leq \eta\mathbb{E}\left[N_H(p)\right]$. We write $G_{\mathcal{L}}$ for a graph drawn from $G(n,p)$ conditioned on $\mathcal{L}_p(H,\eta)$, and use $\mathbb{P}_{\mathcal{L}}$ and $\mathbb{E}_{\mathcal{L}}$ to refer specifically to probability and expectation with respect to this conditional measure.

One notion of similarity we use to characterize the distance between instances of the random graphs is the cut norm. For our purposes, the cut norm can be defined on the set of $n \times n$ matrices as follows:

Definition 4 (Cut norm). For $A \in \mathbb{R}^{n\times n}$,

$$\|A\|_{\square} := \sup_{x,y\in[0,1]^n} \frac{1}{n^2}\left|x^{\top}Ay\right|.$$

We also use the cut norm on vectors $\mathbf{q} \in [0,1]^{\binom{n}{2}}$. In this case we interpret $\|\mathbf{q}\|_{\square} = \|A(\mathbf{q})\|_{\square}$ where $A(\mathbf{q})_{ij} = A(\mathbf{q})_{ji} = q_{i,j}$.

One should think of A as the difference of the adjacency matrices of two (weighted) graphs for this paper. In this case, we obtain the cut norm on graphs alluded to earlier. For a detailed account of the full definition and properties of the cut norm in relation to the theory of graph limits, see the monograph by Lovász (2012).

Many of the ideas in the paper can also be conveniently phrased in the language of hypergraphs, as done in (Kozma and Samotij 2023). A hypergraph will be represented by $\mathcal{H} = (\mathcal{V},\mathcal{E})$. We similarly use $v(\mathcal{H}) = |\mathcal{V}|$ and $e(\mathcal{H}) = |\mathcal{E}|$.

Definition 5 (Hypergraph of copies of H). Let $\mathcal{H}(H)$ be the hypergraph with vertex set $\mathcal{V} = \binom{[n]}{2}$ and edge set $\mathcal{E}$ = copies of H as a subgraph of the complete graph K_n. We will use $\mathcal{H}(H)$ to refer to both the hypergraph and its corresponding edge set.

The following vector representation of a vertex set will be convenient for us to work with.

Definition 6 (Vector representation). Let R_p be a random subset of $\mathcal{V}$ where each vertex is included independently with probability p. Let $Y \in \{0,1\}^{\binom{[n]}{2}}$ be the indicator vector of this random subset conditioned on $\mathcal{L}_p(H,\eta)$, so that Y corresponds to the edges of $G_{\mathcal{L}}$. We write $Y_A = \prod_{e\in A} Y_e$ for the indicator that the subset A is included in R_p.

The relative entropy between random variables plays a key role, as it controls the probability of the lower tail event. Here we define the relevant quantities and list a few useful properties. See (Kozma and Samotij 2023, Section 4) for a more in depth introduction.

Definition 7 (Relative entropy). For a binary vector X define its p-divergence by

$$I_p(X) = D_{\mathrm{KL}}(X \;||\; \mathrm{Ber}(p)^k)$$

where $\mathrm{Ber}(p)^k$ is the k-dimensional product measure with $\mathrm{Ber}(p)$ marginals. We also define the conditional divergence by $I_p(X|Z) = \mathbb{E}\left[I_p(X^Z)\right]$ where X^Z denotes the random variable X conditioned on Z. For any binary random vectors X_1, X_2 and any random variable Z, we have the following useful properties:

1. $I_p(X_1) \geq 0$.
2. $I_p(X_1,X_2) = I_p(X_1|X_2) + I_p(X_2) \geq I_p(X_1) + I_p(X_2)$.
3. $I_p(X_1,X_2|Z) \geq I_p(X_1|Z) + I_p(X_2|Z)$.

We also use the following equality to access the logarithm of the large deviation probability, see, for example, (Kozma and Samotij 2023, Proposition 10) for a proof.

Lemma 2.1 *Let Y be the conditional vector from Definition 6. Then*

$$I_p(Y) = -\log\mathbb{P}\left(\mathcal{L}_p(H,\eta)\right).$$

In the case that X itself is a Bernoulli variable, we recover the relative entropy function. Let $i_p(q) = q \log \frac{q}{p} + (1-q) \log \frac{1-q}{1-p}$ be the relative entropy of q with respect to p. For a vector $\mathbf{q}$, we will write $i_p(\mathbf{q}) := \sum_{e \in \binom{[n]}{2}} i_p(q_e)$ for the total relative entropy with respect to p.

The following variational problem encodes the relationship between entropy and the lower tail event.

Definition 8 (Variational problem). We define the variational problem associated with the lower tail event $\mathcal{L}_p(H, \eta)$ as

$$\Phi_p(H, \eta) := \min_{\mathbf{q} \in [0,1]^{\binom{n}{2}}} \left\{ i_p(\mathbf{q}) : \mathbb{E}\left[N_H(\mathbf{q})\right] \leq \eta \mathbb{E}\left[N_H(p)\right] \right\}.$$

We follow the notation used by Kozma and Samotij (2023), but note here that the relative entropy function is denoted by $I_p(q)$ rather than $i_p(q)$ and the variational problem is denoted by $\mathrm{LT}_p(H, \eta)$ rather than $\Phi_p(H, \eta)$ in (Zhao 2017).

3 | Entropy Increment

In this section, we collect a series of results from the work of Kozma and Samotij (2023) that controls the probability of $\mathcal{L}_p(H, \eta)$ via an entropy increment argument alluded to in the introduction.

The first lemma says that, conditioned on a lower tail event, the marginal probabilities that fixed sets of edges appear can only decrease. The proof is an application of Harris' inequality.

Lemma 3.1 (Kozma and Samotij 2023, Claim 16). *For every $W \subset \mathcal{V}$ and $A \subset \mathcal{V} \setminus W$, $\mathbb{E}\left[Y_A | (Y_w)_{w \in W}\right] \leq p^{|A|}$.*

A key step in (Kozma and Samotij 2023) shows approximate conditional independence of copies of H under the lower tail event by conditioning on a small subset of the edges. This translates to a conditional independence of the appearance of edges of $\mathcal{H}(H)$ given the configuration on a small subset of $\mathcal{V}$. The conditional independence is measured by the following quantities.

$$D_W(B, b) := \mathbb{E}\left[Y_B | (Y_w)_{w \in W}\right] - \mathbb{E}\left[Y_{B \setminus \{b\}} | (Y_w)_{w \in W}\right] \mathbb{E}\left[Y_b | (Y_w)_{w \in W}\right]$$

measures the correlation of a single vertex $b \in \mathcal{V}$ with a hyperedge $B \in \mathcal{E}$ conditioned on the configuration of $W \subset \mathcal{V}$. The weighted sum of these squared correlations is captured by the following quantity:

$$\mathcal{E}_H(W) := \mathbb{E}\left[\sum_{A \in \mathcal{H}(H) - W} \sum_{B \subset A, |B| \geq 2} \sum_{b \in B} \frac{D_W(B, b)^2}{\binom{r}{|B|} |B| p^{2|B|}} \right]. \tag{3.1}$$

Following (Kozma and Samotij 2023), we use $\mathcal{H}(H) - W$ denote the induced sub-hypergraph of $\mathcal{H}(H)$ on the vertex set $\mathcal{V} \setminus W$. It turns out that this quantity controls the difference between a sample from the joint distribution of Y and sampling each edge independently with the corresponding marginal. The proof is an application of inclusion-exclusion and the Cauchy–Schwarz inequality, see (Kozma and Samotij 2023, p. 686) for the proof.

Lemma 3.2. *Let $e(H) = r$ and $W \subset \mathcal{V}$ be a subset of the vertices of $\mathcal{H}(H)$.*

$$\mathbb{E}\left[\sum_{A \in \mathcal{H}(H) - W} \left| \mathbb{E}\left[Y_A | (Y_w)_{w \in W}\right] - \prod_{a \in A} \mathbb{E}\left[Y_a | (Y_w)_{w \in W}\right] \right| \right] \leq p^r ((r-1) e(\mathcal{H}(H) - W))^{1/2} \mathcal{E}_H(W)^{1/2}.$$

Thus, to well-approximate the distribution of (hyper)edges in Y from its independent marginals, we need a W such that both $|W|$ and $\mathcal{E}_H(W)$ are small. The following generalization of (Kozma and Samotij 2023, Lemma 18) finds this small set, and follows from the same proof given in (Kozma and Samotij 2023) and the discussion in (Kozma and Samotij 2023, Section 7).

Lemma 3.3. *Let H be a non-empty graph and $\mathcal{H} = \mathcal{H}(H)$. For all positive α and β, there exists L and V_0 such that the following holds: If $|\mathcal{V}| \geq V_0$ and $p \geq Ln^{-1/m_2(H)}$ and Y is a distribution on subsets of $\mathcal{V}$ such that $\mathbb{E}\left[Y_A | (Y_w)_{w \in W}\right] \leq p^{|A|}$ for all $A \subset \mathcal{V} \setminus W$, there exists a set W with at most $\alpha v(\mathcal{H})$ elements such that $\mathcal{E}_H(W) \leq \beta e(\mathcal{H})$.*

Lemmas 3.1, 3.2, and 3.3 can be combined to obtain the following theorem, showing that the probability of the lower tail event is controlled to the leading term in the exponent by the associated variational problem.

Theorem 3.4 (Kozma and Samotij 2023, Theorem 2). *For every non-empty graph H, $p_0 < 1$, and every $\varepsilon > 0$, there exists a constant L such that for $Ln^{-1/m_2(H)} \leq p \leq p_0$ and every $\eta \in [0, 1]$,*

$$(1-\varepsilon)\Phi_p(H, \eta + \varepsilon) \leq -\log \mathbb{P}\left(\mathcal{L}_p(H, \eta)\right) \leq (1+\varepsilon)\Phi_p(H, \eta(1-\varepsilon)).$$

4 | Stability of the Variational Problem

In the previous section, we related the lower tail probability to an entropy variational problem. The goal of this section is to give a quantitative description of the solution to the variational problem. Our main new technical ingredient is a refinement of (Zhao 2017) that enhances the uniqueness of the solution to $\Phi_p(H, \eta)$ to a stability of the minimizer. In this section, we work with the sparse limit of the variational problem, which is an appropriate substitute for $\Phi_p(H, \eta)$ when n is sufficiently large and $p = o(1)$. We define $h(x) := x \log x - x + 1$ so that $\lim_{p \to 0} p^{-1} i_p(px) = h(x)$ uniformly for $x \in [0, 1]$. It follows that

$$\lim_{p \to 0} p^{-1} \Phi_p(H, \eta) = \Phi(H, \eta)$$

where

$$\Phi(H, \eta) := \min_{\mathbf{q} \in [0,1]^{\binom{n}{2}}} \left\{ \sum_e h(q_e) : \mathbb{E}\left[N_H(\mathbf{q})\right] \leq \eta \cdot N_H(1) \right\}.$$

We know from (Zhao 2017) the following theorem:

Theorem 4.1. *For every graph H there exists an $\eta_H < 1$ such that for all $\eta > \eta_H$, $\Phi(H, \eta)$ is uniquely minimized by the constant function, in this case $\mathbf{q} \equiv \eta^{1/e(H)}$.*

We show that the constant function is stable as the minimizer to $\Phi(H, \eta)$ in this regime in the following quantitative sense.

Proposition 4.2. *Let H be any graph and η_H be as in Theorem 4.1. Then for all $\eta > \eta_H$ there exists a $C_{\eta,H}$ such that the following holds. Let $q = \eta^{1/e(H)}$ and $\varepsilon > 0$ be sufficiently small in terms of η and H. Then $\mathbf{q} \equiv q$ is the unique minimizer of $\Phi(H, \eta)$, and if $\sum_e h(q_e) < \Phi(H, \eta) + \varepsilon^{12}\binom{n}{2}$, then $\|\mathbf{q} - q\|_\square < C_{\eta,H}\varepsilon$.*

To begin, we first give a weak description of a near minimizer, showing that only a few of its entries can be very small. We will use subgraph count density notation in this section, writing $t_{\text{inj}}(H, \mathbf{q}) = \frac{\mathbb{E}[N_H(\mathbf{q})]}{N_H(1)}$. This is also known as the injective homomorphism density, closely related to the standard notion of homomorphism density (see Zhao 2023, Section 4.3 for more on homomorphism densities). The following is the stability version of (Zhao 2017, Lemma 5.2).

Lemma 4.3. *Let $\eta > \eta_H$ as in Theorem 4.1. For all ε sufficiently small in terms of η, H, if $\sum_e h(q_e) \leq \binom{n}{2} h(\eta^{1/e(H)}) + \varepsilon^2 \binom{n}{2}$ and $t_{\text{inj}}(H, \mathbf{q}) < \eta$ then $q_e \geq c := \eta^{1/\eta}$ for all but $\varepsilon\binom{n}{2}$ edges e.*

Proof. The idea is to show that if the set of small entries is too large in a near minimizer, then a marginal boost of the edge probabilities on this set leads to a configuration with few copies of H that has too small of an entropy.

Let $B = \{e : q_e < c\}$ denote the bad set of edges and let $\delta = \frac{|B|}{\binom{n}{2}}$. Suppose we increase $\mathbf{q}$ on each of these edges to obtain $\mathbf{q}' := \mathbf{q} + \gamma \mathbb{1}_B$. Since h is convex and decreasing, for every $e \in B$ we have

$$h(q_e + \gamma) - h(q_e) \leq h(c + \gamma) - h(c) = -\gamma(1 - \log(c + \gamma)) - c \log(1 - \frac{\gamma}{c + \gamma}) < 0.$$

Setting $q := \eta^{1/e(H)}$, we can bound the new entropy by

$$\begin{aligned}\sum_e h(q'_e) &= \sum_{e\in B} h(q_e+\gamma) + \sum_{e\in B^c} h(q_e)\\ &\le \sum_e h(q_e) + |B|\left(-\gamma(1-\log(c+\gamma)) - c\log(1-\frac{\gamma}{c+\gamma})\right)\\ &\le \binom{n}{2}\left(h(q)+\epsilon^2 - \delta\gamma(1-\log(c+\gamma)) - \delta c\log(1-\frac{\gamma}{c+\gamma})\right).\end{aligned}$$

On the other hand, since $\|\mathbf{q}'-\mathbf{q}\|_\square = \|\gamma \mathbb{1}_B\|_\square < \gamma\delta$, the counting lemma (see, e.g., Zhao 2023, Section 4.5) implies

$$t_{\text{inj}}(H,\mathbf{q}') - t_{\text{inj}}(H,\mathbf{q}) \le e(H)\|\mathbf{q}'-\mathbf{q}\|_\square \le \gamma\delta e(H).$$

In particular,

$$t_{\text{inj}}(H,\mathbf{q}') \le \eta + \gamma\delta e(H).$$

By Theorem 4.1 we know that the variational problem $\Phi(H, \eta+\gamma\delta e(H))$ is minimized by the constant function $(\eta+\gamma\delta e(H))^{1/e(H)}$, implying that

$$\begin{aligned}\sum_e h(q'_e) &\ge \binom{n}{2} h((\eta+\gamma\delta e(H))^{1/e(H)}) \ge \binom{n}{2} h(q+\gamma\delta\eta^{-1})\\ &= \binom{n}{2}\left(h(q) - \frac{\gamma\delta}{\eta}\left(1-\log(q+\frac{\gamma\delta}{\eta})\right) - q\log(1-\frac{\gamma\delta\eta^{-1}}{q+\gamma\delta\eta^{-1}})\right).\end{aligned}$$

Combining the two bounds on $\sum_e h(q'_e)$, we must have

$$-\frac{\gamma\delta}{\eta}\left(1-\log(q+\frac{\gamma\delta}{\eta})\right) - q\log(1-\frac{\gamma\delta\eta^{-1}}{q+\gamma\delta\eta^{-1}}) \le \epsilon^2 - \delta\gamma(1-\log(c+\gamma)) - \delta c\log(1-\frac{\gamma}{c+\gamma}).$$

By Taylor expansion, we know

$$\delta c\log(1-\frac{\gamma}{c+\gamma}) = -\frac{\delta c\gamma}{c+\gamma} - O_\eta(\delta\gamma^2) = -\gamma\delta - O_\eta(\delta\gamma^2)$$

and similarly

$$q\log(1-\frac{\gamma\delta\eta^{-1}}{q+\gamma\delta\eta^{-1}}) = -\frac{\gamma\delta\eta^{-1}}{1+\gamma\delta(q\eta)^{-1}} - O_\eta(\delta^2\gamma^2) = -\gamma\delta\eta^{-1} - O_\eta(\delta^2\gamma^2).$$

Rearranging, we obtain

$$\begin{aligned}\gamma\delta\left(1-\log(c+\gamma) - \frac{1}{\eta} + \frac{1}{\eta}\log(q+\frac{\gamma\delta}{\eta}) - 1 + \frac{1}{\eta}\right) + O_\eta(\delta\gamma^2) &\le \epsilon^2\\ \gamma\delta\left(-\log c + \frac{1}{\eta}\log q\right) + O_\eta(\delta\gamma^2) &\le \epsilon^2.\end{aligned}$$

Setting $\gamma = \frac{2\epsilon}{\frac{\log q}{\eta} - \log c}$, we deduce that for all ϵ sufficiently small in terms of η and H, the number of bad edges is $\le \epsilon\binom{n}{2}$. □

Before proving the stability, we record a useful fact that appears as (Zhao 2017, Fact 5.5).

Fact 4.4. Let η_H be the unique solution in the interval $(0,1)$ to the equation

$$h(\eta^{1/\eta}) = h(\eta^{1/e(H)}) + \eta^{1/e(H)}\log(\eta^{1/e(H)})(\log(\eta^{1/\eta}) - \log(\eta^{1/e(H)}))$$

Then for all $(x,r) \in [\eta_H^{1/\eta_H}, 1] \times [\eta_H^{1/e(H)}, 1]$ the inequality $h(x) \ge h(r) + r\log(r)(\log x - \log r)$ holds.

Proof of Proposition 4.2. The strategy is as follows. We use Lemma 4.3 and Fact 4.4 to give a lower bound on the average value of $\log q_e$. Then, a nearly matching upper bound follows by Jensen's inequality. The combination of these two bounds implies that each inequality must be nearly tight, from which we can extract term-by-term bounds on q_e.

Applying Lemma 4.3 we know that $B = \{e : q_e < \eta^{1/\eta}\}$ has size at most $\varepsilon^6\binom{n}{2}$. We show that for most of the remaining edges, q_e must be close to $q := \eta^{1/e(H)}$. Using Fact 4.4, we lower bound

$$\sum_e h(q_e) \geq \sum_{e\in B^c} h(q_e) \geq \sum_{e\in B^c} \left(h(q) + q\log q(\log q_e - \log q)\right) \tag{4.1}$$

$$= |B^c|h(q) + q\log q\left(\sum_{e\in B^c}(\log q_e - \log q)\right). \tag{4.2}$$

Combining this with the assumption that q_e is a near minimizer, we obtain

$$\sum_{e\in B^c}(\log q - \log q_e) \leq \frac{\varepsilon^6 + \varepsilon^{12}}{q\log\frac{1}{q}}\binom{n}{2}$$

so that

$$\frac{1}{\binom{n}{2}}\sum_{e\in B^c}\log q_e \geq (1-\varepsilon^6)\log q - \frac{\varepsilon^6+\varepsilon^{12}}{q\log\frac{1}{q}} = \log q - O_{\eta,H}(\varepsilon^6). \tag{4.3}$$

We now work out the upper bound. Recall that $N_H(1)$ is the total number of copies of H in K_n. By symmetry, each edge is in $\frac{e(H)}{\binom{n}{2}}N_H(1)$ copies of H. Thus, the number of copies of H that intersect B is at most $\varepsilon^6 e(H)N_H(1)$. Let X be a uniformly random copy of H that is contained in B^c. Then

$$\begin{aligned}\mathbb{E}\left[\log\left(\prod_{e\in X}q_e\right)\right] &= \frac{1}{N_H(\mathbb{1}_{B^c})}\sum_H\sum_{e\in H}\log q_e = \frac{1}{N_H(\mathbb{1}_{B^c})}\sum_{e\in B^c}\log q_e\cdot|\{H : e\in H\}| \\ &\geq \frac{1}{N_H(\mathbb{1}_{B^c})}\sum_{e\in B^c}\frac{e(H)}{\binom{n}{2}}N_H(1)\log q_e \geq \frac{e(H)}{\binom{n}{2}(1-\varepsilon^6 e(H))}\sum_{e\in B^c}\log q_e.\end{aligned}$$

On the other hand, Jensen's inequality implies

$$\mathbb{E}\left[\log\left(\prod_{e\in X}q_e\right)\right] \leq \log\mathbb{E}\left[\prod_{e\in X}q_e\right] \leq \log\left(\frac{\eta}{1-\varepsilon^6 e(H)}\right) = e(H)\log q - \log(1-\varepsilon^6 e(H)).$$

Combining these inequalities yields

$$\frac{1}{\binom{n}{2}}\sum_{e\in B^c}\log q_e \leq (1-\varepsilon^6 e(H))\log q - \frac{1-\varepsilon^6 e(H)}{e(H)}\log(1-\varepsilon^6 e(H)) = \log q + O_{\eta,H}(\varepsilon^6). \tag{4.4}$$

The lower and upper bounds (4.3) and (4.4) imply that the term-wise application of Fact 4.4 in Equation (4.1) must be nearly tight. Quantitatively, the number of edges e for which

$$h(q_e) > h(q) + q\log q(\log q_e - \log q) + \varepsilon^3$$

is at most $O_{\eta,H}(\varepsilon^3)\binom{n}{2}$. One can check that the equation $h(q_e) = h(q) + q\log q(\log q_e - \log q)$ has at most two solutions on $[0,1]$, one of which is $q_e = q$.

We analyze the order of vanishing of the function $h(q_e) - h(q) - q\log q(\log q_e - \log q)$ at each of the roots. Taking two derivatives, we get the expression $\frac{1}{q_e} + \frac{q\log q}{q_e^2}$, which vanishes at $q_e = q\log\frac{1}{q}$. On the other hand, the third derivative is

$-\frac{1}{q_e^2} - \frac{2q\log q}{q_e^3}$, which vanishes at $q_e = 2q\log\frac{1}{q}$. In particular, since $0 < q < 1$, one of these two derivatives does not vanish at every point q_e. This implies that the order of vanishing at each of the roots is at most 3.

Thus, for ε sufficiently small in terms of η and H, if $h(q_e) \le h(q) + q\log q(\log q_e - \log q) + \varepsilon^3$ we must have that q_e is within $O(\varepsilon)$ of one of these solutions. Finally, since we know that the average value of $\log q_e$ is within $O(\varepsilon^6)$ of $\log q$, all but an $O(\varepsilon^6)$ fraction of the q_e must in fact be close to q rather than the second root. This shows that at most $O(\varepsilon^3\binom{n}{2})$ of the edges q_e have $|q_e - q| > \varepsilon$, which implies that $\|\mathbf{q} - q\|_\square < C_{\eta,H}\varepsilon$. □

5 | Proof of the Main Theorem

In this section, we prove the main results, Theorem 1.1 and Corollary 1.2. For the rest of the section, we retain the following setup: Fix a graph H' and suppose $\eta > \eta_{H'}$. Let $G_{\mathcal{L}} \sim G(n,p)$ conditioned on $\mathcal{L}_p(H',\eta)$. Let $q = \eta^{1/e(H')}p$ be the unique minimizer of $\Phi_p(H',\eta)$. We will use the notation $x \lesssim y$ to mean that there exists a constant C such that $x \le Cy$.

The following lemma establishes a weak notion of typical structure by showing that the conditional distribution can be decomposed into a mixture of a small number of measures, almost all of which have marginals that are close to the optimizer of the variational problem.

Lemma 5.1. *For any $\varepsilon > 0$ there exists an L such that if $p \ge Ln^{-1/m_2(H')}$ then the following holds. There exists a set of edges W_0 such that for any $W \supseteq W_0$ and $|W| < \frac{\varepsilon}{e(H')}\binom{n}{2}$ we have*

$$\mathbb{P}\left(\left\|\mathbf{q}^W - q\right\|_\square > \varepsilon p\right) \lesssim \varepsilon$$

where $q_e^W = \mathbb{E}\left[Y_e|(Y_w)_{w\in W}\right]$ for $e \notin W$ and $q_e^W = p$ otherwise.

Proof. Throughout this proof, we may assume that ε is sufficiently small as a function of η and H'. Set $\alpha = \frac{\varepsilon}{2e(H')}$ and $\beta = \frac{\varepsilon^4}{e(H')}$. Let L be large enough to invoke Lemma 3.3 on $\mathcal{H}(H')$. We obtain a set W_0 such that $|W_0| \le \frac{\varepsilon}{2e(H')}\binom{n}{2}$ and $\mathcal{E}_{H'}(W_0) \le \frac{\varepsilon^4}{e(H')}\binom{n}{v(H')}$ where $\mathcal{E}_{H'}$ is defined as in Equation (3.1). Now let $W \supseteq W_0$ be any set with $\le \frac{\varepsilon}{e(H')}\binom{n}{2}$ edges. Note that $\mathcal{E}_{H'}$ is a decreasing function with respect to containment, so $\mathcal{E}_{H'}(W) \le \frac{\varepsilon^4}{e(H')}\binom{n}{v(H')}$ as well. Applying Lemma 3.2, we obtain

$$\begin{aligned}\mathbb{E}\left[\sum_{A\in\mathcal{H}(H')-W}\left|\mathbb{E}\left[Y_A|(Y_w)_{w\in W}\right] - \prod_{a\in A}\mathbb{E}\left[Y_a|(Y_w)_{w\in W}\right]\right|\right] &\le p^{e(H')}(e(H')e(\mathcal{H}(H')-W))^{1/2}\mathcal{E}_{H'}(W)^{1/2}\\ &\le \varepsilon^2 p^{e(H')}\binom{n}{v(H')}.\end{aligned}$$

By Markov's inequality, we have that with probability at least $1-\varepsilon$,

$$\sum_{A\in\mathcal{H}(H')-W}\prod_{a\in A}\mathbb{E}\left[Y_a|(Y_w)_{w\in W}\right] \le \sum_{A\in\mathcal{H}(H')-W}\mathbb{E}\left[Y_A|(Y_w)_{w\in W}\right] + \varepsilon p^{e(H')}\binom{n}{v(H')}.$$

Since Y is drawn from the lower tail distribution, it holds with probability 1 that

$$\sum_{A\in\mathcal{H}(H')-W}\mathbb{E}\left[Y_A|(Y_w)_{w\in W}\right] \le \eta p^{e(H')}\binom{n}{v(H')}$$

and so with probability $1-\varepsilon$, we have

$$\sum_{A\in\mathcal{H}(H')-W}\prod_{a\in A}\mathbb{E}\left[Y_a|(Y_w)_{w\in W}\right] \le (\eta+\varepsilon)p^{e(H')}\binom{n}{v(H')}.$$

Define $q_e^W := \mathbb{E}\left[Y_e | (Y_w)_{w\in W}\right]$ for $e \notin W$ and $q_e^W = p$ for $e \in W$. The size of W along with symmetry of edges in $\mathcal{H}(H')$ implies that there are at most $\varepsilon n^{v(H')}$ edges $A \in \mathcal{H}(H')$ such that $A \cap W \neq \emptyset$. Lemma 3.1 implies that $q_e^W \le p$ for every vertex e. These two estimates combined ensure that

$$\sum_{\substack{A\in\mathcal{H}(H')\\ A\cap W\neq\emptyset}} \prod_{a\in A} q_a^W \le \varepsilon p^{e(H')}\binom{n}{v(H')}$$

and so altogether we have

$$\sum_{A\in\mathcal{H}(H')} \prod_{a\in A} q_a^W \le (\eta + 2\varepsilon) p^{e(H')}\binom{n}{v(H')}.$$

In particular, $\mathbf{q}^W$ satisfies $\mathcal{L}_p(H', \eta + 2\varepsilon)$ with probability at least $1 - \varepsilon$.

We use this fact to argue that $\mathbf{q}^W$ must be close to the constant vector q, where $q = \eta^{1/e(H')} p$. Since $\mathbf{q}^W$ satisfies $\mathcal{L}_p(H', \eta + 2\varepsilon)$ with probability at least $1 - \varepsilon$, it must hold with probability at least $1 - \varepsilon$ that

$$i_p(\mathbf{q}^W) \ge \Phi_p(H', \eta + 2\varepsilon).$$

Let $a(\delta) := \mathbb{P}\left(i_p(\mathbf{q}^W) > \Phi_p(H', \eta) + \delta\binom{n}{2}p\right)$. This gives a lower bound of

$$a(\delta)(\Phi_p(H', \eta) + \delta\binom{n}{2}p) + (1 - \varepsilon - a(\delta))\Phi_p(H', \eta + 2\varepsilon) \le \mathbb{E}\left[i_p(\mathbf{q}^W)\right].$$

From Theorem 3.4, we can deduce an upper bound on $\mathbb{E}\left[i_p(\mathbf{q}^W)\right]$ via relative entropy manipulations. Using the properties from Definition 7 and Lemma 2.1,

$$\begin{aligned}\mathbb{E}\left[i_p(\mathbf{q}^W)\right] &= \sum_{e\in\mathcal{V}\setminus W} I_p(Y_e | (Y_w)_{w\in W}) \le I_p((Y_e)_{e\in\mathcal{V}\setminus W} | (Y_w)_{w\in W}) + I_p((Y_w)_{w\in W}) \\ &\le I_p(Y) = -\log \mathbb{P}\left(\mathcal{L}_p(H', \eta)\right) \le (1+\varepsilon)\Phi_p(H', \eta(1-\varepsilon)).\end{aligned}$$

Combining the two inequalities and rearranging for $a(\delta)$,

$$a(\delta) \le \frac{(1+\varepsilon)\Phi_p(H', \eta(1-\varepsilon)) - (1-\varepsilon)\Phi_p(H', \eta + 2\varepsilon)}{\Phi_p(H', \eta) + \delta\binom{n}{2}p - \Phi_p(H', \eta + 2\varepsilon)} \lesssim \frac{\varepsilon}{\varepsilon + \delta}.$$

Here we used the fact that each of the variational problems in the expression is minimized by the constant graphon, and the optimal value is locally Lipschitz in η. By Proposition 4.2,

$$\mathbb{P}\left(\left\|\mathbf{q}^W - q\right\|_\square > \delta p\right) \le \alpha\left(\left(\frac{\delta}{C_{\eta,H}}\right)^{12}\right) \lesssim \frac{\varepsilon}{\varepsilon + \delta^{12}}.$$

Setting $\delta = \varepsilon^{1/13}$ finishes the proof. □

To boost this weak structure to a typical structure in terms of subgraph counts, we first estimate the expected count of any other graph H. This is done by once again invoking the entropy increment to gain control over copies of H under the conditional distribution.

Theorem 5.2. *Fix a graph H and let $m = \max\{m_2(H), m_2(H')\}$. Then for any $\varepsilon > 0$ there exists an L such that if $p \ge Ln^{-1/m}$,*

$$\left|\mathbb{E}_{\mathcal{L}}[N_H(G_{\mathcal{L}})] - \mathbb{E}\left[N_H(q)\right]\right| < \varepsilon\mathbb{E}\left[N_H(q)\right].$$

Proof. Once again, we may assume that ε is sufficiently small as a function of η, H', and H. Let L be large enough so that the conditions of Lemma 3.3 are satisfied for both H and H', and the conditions of Lemma 5.1 are satisfied for H'. Let W_0 be the set guaranteed by Lemma 5.1 with parameter $\frac{\varepsilon}{2e(H)e(H')}$. Invoking Lemma 3.3 on $\mathcal{H}(H)$ with $\alpha = \frac{\varepsilon}{2e(H)e(H')}$ and

$\beta = \frac{\varepsilon^4}{e(H)}$, we obtain a set W_1 such that $|W_1| \leq \frac{\varepsilon}{2e(H)e(H')}\binom{n}{2}$ and $\mathcal{E}_H(W_1) \leq \frac{\varepsilon^4}{e(H)}\binom{n}{v(H)}$. Let $W = W_0 \cup W_1$, then $|W| \leq \frac{\varepsilon}{e(H)e(H')}$ and $\mathcal{E}_H(W) \leq \frac{\varepsilon^4}{e(H)}\binom{n}{v(H)}$. Moreover, by choice of W_0 and Lemma 5.1 we also have that

$$\mathbb{P}\left(\left\|\mathbf{q}^W - q\right\|_\square > \varepsilon p\right) \leq C_{\eta,H'}\varepsilon.$$

Note that since $\mathbf{q}^W \leq p$ pointwise by Lemma 3.1, the graphons $\frac{\mathbf{q}^W}{p}$ and $\frac{q}{p}$ are bounded above by 1. Thus, the counting lemma (Zhao 2023, Section 4.5) implies that

$$\left|\mathbb{E}\left[N_H(\mathbf{q}^W)\right] - \mathbb{E}\left[N_H(q)\right]\right| = p^{e(H)}\left|\mathbb{E}\left[N_H\left(\frac{\mathbf{q}^W}{p}\right)\right] - \mathbb{E}\left[N_H\left(\frac{q}{p}\right)\right]\right| \leq p^{e(H)}n^{v(H)}e(H)\left\|\frac{\mathbf{q}^W}{p} - \frac{q}{p}\right\|_\square.$$

Here we treat $\mathbf{q}^W$ as fixed and the expectation is taken over the randomness of $N_H(G)$ where G is a random graph sampled from $G(n, \mathbf{q}^W)$. This implies that

$$\mathbb{P}\left(\left|\mathbb{E}\left[N_H(\mathbf{q}^W)\right] - \mathbb{E}\left[N_H(q)\right]\right| > \varepsilon e(H)p^{e(H)}n^{v(H)}\right) \leq \mathbb{P}\left(\left\|\mathbf{q}^W - q\right\|_\square > \varepsilon p\right) \leq C_{\eta,H'} \cdot \varepsilon.$$

Now we control the expected number of copies of H. We decompose

$$\mathbb{E}\left[N_H(Y)\right] = \mathbb{E}\left[\sum_{A\in\mathcal{H}(H)} Y_A\right] = \mathbb{E}\left[\sum_{A\in\mathcal{H}(H)-W} Y_A\right] + \mathbb{E}\left[\sum_{A\in\mathcal{H}(H), A\cap W\neq\emptyset} Y_A\right].$$

For the second expectation, note that the size of W along with symmetry of the edges in $\mathcal{H}(H)$ implies that there are at most $\varepsilon n^{v(H)}$ terms in the sum. Moreover, Lemma 3.1 implies that $\mathbb{E}\left[Y_A\right] \leq p^{e(H)}$ for every A, and thus the contribution of this term is bounded by $\varepsilon p^{e(H)}n^{v(H)}$.

For the first expectation, applying Lemma 3.2 and an analogous computation to the proof of Lemma 5.1 to the hypergraph $\mathcal{H}(H)$, we have that with probability at least $1 - \varepsilon$,

$$\sum_{A\in\mathcal{H}(H)-W}\left|\mathbb{E}\left[Y_A|(Y_w)_{w\in W}\right] - \prod_{a\in A}q_a^W\right| \leq \varepsilon p^{e(H)}\binom{n}{v(H)}.$$

Moreover, recall that with probability at least $1 - C_{\eta,H'}\varepsilon$ we have

$$\left|\sum_{A\in\mathcal{H}(H)}\left(\prod_{a\in A}q_a^W - q^{e(H)}\right)\right| < \varepsilon e(H)p^{e(H)}n^{v(H)}.$$

Once again, since $\mathbf{q}^W$ and q are both bounded by p and W is small, we have

$$\begin{aligned}\left|\sum_{A\in\mathcal{H}(H)-W}\left(\prod_{a\in A}q_a^W - q^{e(H)}\right)\right| &\leq \left|\sum_{A\in\mathcal{H}(H)}\left(\prod_{a\in A}q_a^W - q^{e(H)}\right)\right| + \left|\sum_{A\in\mathcal{H}(H), A\cap W\neq\emptyset}\left(\prod_{a\in A}q_a^W - q^{e(H)}\right)\right| \\ &< 2\varepsilon e(H)p^{e(H)}n^{v(H)}\end{aligned}$$

with probability at least $1 - C_{\eta,H'}\varepsilon$. Thus, with probability at least $1 - C_{\eta,H'}\varepsilon$,

$$\left|\sum_{A\in\mathcal{H}(H)-W}\left(\mathbb{E}\left[Y_A|(Y_w)_{w\in W}\right] - q^{e(H)}\right)\right| \leq 3\varepsilon e(H)p^{e(H)}n^{v(H)}.$$

Note that by Lemma 3.1 we know that the random variable $\mathbb{E}\left[\sum_{A\in\mathcal{H}(H)-W} Y_A\middle|(Y_w)_{w\in W}\right]$ is uniformly bounded above by $|\mathcal{H}(H) - W|p^{e(H)} \leq n^{v(H)}p^{e(H)}$. Thus,

$$\left|\mathbb{E}\left[\sum_{A\in\mathcal{H}(H)-W} Y_A\right]-\binom{n}{v(H)}q^{e(H)}\right|=\left|\mathbb{E}\left[\mathbb{E}\left[\sum_{A\in\mathcal{H}(H)-W} Y_A\middle|(Y_w)_{w\in W}\right]-\binom{n}{v(H)}q^{e(H)}\right]\right|$$
$$\lesssim_{\eta,H,H'}\varepsilon p^{e(H)}n^{v(H)}.$$

Combining the estimates together, we obtain that

$$\left|\mathbb{E}\left[N_H(Y)\right]-\binom{n}{v(H)}q^{e(H)}\right|\lesssim_{\eta,H,H'}\varepsilon p^{e(H)}n^{v(H)}$$

which is the desired conclusion. □

Finally, to prove Theorem 1.1, we further boost the estimate on the expected number of copies of H to a concentration of the number of copies of H by a second moment argument. The key observation is that the variance can be written in terms of expected subgraph counts, which are controlled by Theorem 5.2.

Proof of Theorem 1.1. We show that the subgraph count H concentrates well around its expectation, which by Theorem 5.2 implies that it is close to $\mathbb{E}\left[N_H(q)\right]$ as well.

$$\operatorname{Var}\left(N_H(Y)\right)=\operatorname{Var}\left(\sum_{A\in\mathcal{H}(H)} Y_A\right)=\sum_{A,A'\in\mathcal{H}(H)}\mathbb{E}\left[Y_AY_{A'}\right]-\mathbb{E}\left[Y_A\right]\mathbb{E}\left[Y'_A\right].$$

We partition this sum into cases based on the graph Γ induced by the edge set $A\cup A'$. Suppose first that $A\cap A'\neq\emptyset$. Denote the graph of their intersection by Γ', and note that Γ' is a subgraph of H. By Lemma 3.1 we can bound

$$\mathbb{E}\left[Y_AY_{A'}\right]\le p^{e(\Gamma)}=p^{2e(H)-|e(\Gamma')|}.$$

Moreover, there are at most $n^{v(\Gamma)}=n^{2v(H)-v(\Gamma')}$ pairs of A and A' that induce this graph. Thus, the total contribution is bounded by

$$\frac{n^{2v(H)}p^{2e(H)}}{n^{v(\Gamma')}p^{e(\Gamma')}}\ll n^{2v(H)}p^{2e(H)}$$

for the given regime of p. Thus, it only remains to consider $A\cap A'=\emptyset$, in other words, $\Gamma=$ the disjoint union of two copies of H. Note that $m_2(\Gamma)=m_2(H)$ and thus by Theorem 5.2 we know that for L large enough $\left|\mathbb{E}\left[N_\Gamma(Y)\right]-\mathbb{E}\left[N_\Gamma(q)\right]\right|<\varepsilon^3p^{e(\Gamma)}n^{v(\Gamma)}$. By the symmetry of the copies of Γ, this implies that $\left|\mathbb{E}\left[Y_AY_{A'}\right]-q^{e(\Gamma)}\right|<\varepsilon^3p^{e(\Gamma)}$. In particular, for the pairs $A\cap A'=\emptyset$ the contribution is bounded by $\varepsilon^3p^{e(\Gamma)}$. There are at most $n^{2v(H)}$ pairs in the sum, so

$$\operatorname{Var}\left(N_H(Y)\right)\le\varepsilon^3p^{2e(H)}n^{2v(H)}\le\varepsilon^3\mathbb{E}\left[N_H(Y)\right]^2.$$

The result follows from Chebyshev's inequality.

To deduce Corollary 1.2, we show that there exist finite graphs that certify pseudorandomness with 2-density arbitrarily close to 1. As a consequence, typicality in the sense of cut norm follows from typicality in the sense of subgraph counts.

Proof of Corollary 1.2. Choose k large enough such that $\frac{k}{k-1}<m_2(H)$. Since $m_2(C_{2k})=\frac{2k-1}{2k-2}\le\frac{k}{k-1}<m_2(H)$, we may apply Theorem 1.1 to H and C_{2k} and obtain some constant L such that for $p\ge Ln^{-1/m_2(H)}$, $G_{\mathcal{L}}$ has the correct C_{2k} count with probability at least $1-\varepsilon$. We show that C_{2k} is forcing for all $p\ge Ln^{-\frac{1}{m_2(H)}}\gg n^{-\frac{k-1}{k}}$. Forcing refers to the property that having the correct C_{2k} count relative to the density of a graph implies pseudorandomness (see Zhao 2023, Chapter 3 for further discussion of forcing graphs). This implies that $G_{\mathcal{L}}$ is close to the constant graphon with probability at least $1-\varepsilon$.

Let A be the adjacency matrix of $G_{\mathcal{L}}$. We know that with probability at least $1-\varepsilon$,

$$\left|N_{C_{2k}}(G_{\mathcal{L}})-\binom{n}{2k}q^{2k}\right|<\varepsilon n^{2k}p^{2k}.$$

Note that $\mathrm{Tr}(A^{2k})$ counts the number of closed walks of length $2k$ in the graph $G_{\mathcal{L}}$. Let H be any graph that can be a closed walk of length $2k$. By Harris's inequality $\mathbb{E}\left[N_H(G_{\mathcal{L}})\right] \le \binom{n}{v(H)} p^{e(H)}$. This is much smaller than $n^{2k}p^{2k}$ as soon as $p \gg n^{-\frac{2k-v(H)}{2k-e(H)}}$. This quantity is maximized when H is a tree with k edges (each of which is traversed twice to form a closed walk). Since $p \gg n^{-\frac{k-1}{k}}$, for any such graph H the expected contribution is $o(n^{2k}p^{2k})$, and thus with probability $1-o(1)$ $N_H(G_{\mathcal{L}}) = o(n^{2k}p^{2k})$. It remains to show that $\|A - qJ\|_{\square} < \varepsilon p$ where J is the all 1's matrix. The above computation verifies that $\left|\mathrm{Tr}((A-qJ)^{2k})\right| < 2\varepsilon n^{2k}p^{2k}$ with probability $1-2\varepsilon$ and on this event the eigenvalues of $A - qJ$ are all at most $\varepsilon^{1/2k}np$. For any vectors $x, y \in [0,1]^n$, write $x = \sum_{i=1}^n a_i v_i$ and $y = \sum_{i=1}^n b_i v_i$ where $\{v_i\}$ is an orthonormal basis of eigenvectors of $A - qJ$. Then

$$\begin{aligned}\left|x^\top (A-qJ)y\right| &= \left|\sum_{i=1}^n a_i b_i \lambda_i\right| \le (\max_i |\lambda_i|)\sum_{i=1}^n |a_i b_i| \le o(np)\cdot \left(\sum_{i=1}^n a_i^2\right)^{1/2}\left(\sum_{i=1}^n b_i^2\right)^{1/2} \\ &= \varepsilon^{1/2k} np \cdot \|x\|_2 \|y\|_2 \le \varepsilon^{1/2k} n^2 p.\end{aligned}$$

The conclusion follows since $\left\|G_{\mathcal{L}} - q\right\|_{\square} = \sup_{x,y\in[0,1]^n} \frac{1}{n^2}\left|x^\top(A-qJ)y\right| = \varepsilon^{1/2k}p$ and $\varepsilon > 0$ was chosen arbitrarily. □

6 | Discussion: Extension to General Hypergraphs

To conclude, we discuss the applicability of this argument in the more general hypergraph setup of (Kozma and Samotij 2023). We first introduce the notation that extends our study of lower tails in hypergraphs encoding subgraphs—namely $\mathcal{H}(H)$—to general hypergraphs. Recall that we denote a hypergraph by $\mathcal{H} = (\mathcal{V}, \mathcal{E})$. First, we define the corresponding lower tail event, which says that a random subset of the vertices contains fewer edges than expected.

Definition 9 (General lower tail event). For $\mathbf{q} \in [0,1]^{\mathcal{V}}$ let $\mathcal{V}_{\mathbf{q}}$ be a random subset of $\mathcal{V}$ where each vertex v is retained independently with probability q_v. When $q_v = p$ for every v we write $\mathcal{V}_p$. The lower tail event $\mathcal{L}_p(\mathcal{H}, \eta)$ is the event that $e(\mathcal{H}[U]) \le \eta\mathbb{E}\left[e(\mathcal{H}[\mathcal{V}_p])\right]$. Here $\mathcal{H}[U]$ denotes the restriction of the hypergraph $\mathcal{H}$ to the vertex subset $U \subset \mathcal{V}$.

We can analogously define the variational problem associated with this lower tail event.

Definition 10 (General variational problem). The variational problem associated with the lower tail event $\mathcal{L}_p(\mathcal{H}, \eta)$ is

$$\Phi_p(\mathcal{H}, \eta) := \min_{\mathbf{q}\in[0,1]^{\mathcal{V}}} \{ i_p(\mathbf{q}) : \mathbb{E}\left[e(\mathcal{H}[\mathcal{V}_{\mathbf{q}}])\right] \le \eta\mathbb{E}\left[e(\mathcal{H}[\mathcal{V}_p])\right] \}.$$

All of the results of (Kozma and Samotij 2023) that we recounted in Section 3 have analogs that hold for hypergraphs satisfying the following uniformity of degrees condition. In particular, Kozma and Samotij show that Theorem 3.4 holds for hypergraphs satisfying Condition 1—that the logarithm of the lower tail probability is controlled by the associated variational problem.

Condition 1. For an r-uniform hypergraph $\mathcal{H}$, $v(\mathcal{H}) = |V(\mathcal{H}|$, $e(\mathcal{H}) = |E(\mathcal{H})|$, and $\deg_{\mathcal{H}}(B) = |\{e \in E(\mathcal{H}) : B \subset e\}|$. Let $\Delta_s := \max\{\deg_{\mathcal{H}}(B) : |B| = s\}$ be the maximum degree of a set of s vertices. Then there exists λ and K such that the following holds for every $s \in [r]$:

$$\Delta_s(\mathcal{H}) \le K(\lambda p)^{s-1}\frac{e(\mathcal{H})}{v(\mathcal{H})}.$$

The condition is closely related to that of the hypergraph container method, introduced independently by Balogh et al. (2015) and Saxton and Thomason (2015), see (Balogh et al. 2018) for a survey of this powerful new technique. For further discussion of this connection, see the introduction of (Kozma and Samotij 2023).

In this general setup, our argument extends in the following manner. Suppose we have two hypergraphs $\mathcal{H}$ and $\mathcal{H}'$ that live on the same vertex set and satisfy Condition 1 under the same parameters. We can apply the entropy increment to decompose the lower tail measure with respect to both $\mathcal{H}$ and $\mathcal{H}'$ simultaneously. If we know that $\Phi_p(\mathcal{H}', \eta)$ has a stability of the minimizing vector $\mathbf{q}$, we can deduce that the expected number of edges of $\mathcal{H}$ in a random sample of vertices from $\mathcal{V}_p$ conditioned on $\mathcal{L}_p(\mathcal{H}', p)$ is close to the expectation under $\mathbf{q}$ (analogous to Theorem 5.2). To boost this to a concentration statement, we need to understand the pairwise interactions between edges $\mathcal{H}$. Under the further assumption that pairs of intersecting edges in $\mathcal{H}$ also satisfy the uniformity Condition 1, we obtain that the number of edges of $\mathcal{H}$ in a random sample of vertices from the conditional measure is close to the expectation under $\mathbf{q}$ with high probability (analogous to Theorem 1.1). This reduces the problem of counting edges sampled from a lower tail measure to analyzing the variational problem and the uniformity of pairs of edges in the hypergraph.

The argument is also robust to conditioning on multiple lower tail events simultaneously, as well as counting edges of multiple hypergraphs simultaneously. However, given the difficulty of resolving the variational problem already for a single lower tail, applications of this form seem farther from fruition.

Acknowledgments

The author thanks Yufei Zhao for introducing the problem and for many conversations about this work. The author also thanks Wojciech Samotij and Mehtaab Sawhney for helpful discussions and anonymous referees for detailed feedback that improved the paper. The author is supported by NSF GRFP 2141064. Open Access funding enabled and organized by Massachusetts Institute of Technology

Data Availability Statement

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

References

Balogh, J., R. Morris, and W. Samotij. 2015. "Independent Sets in Hypergraphs." *Journal of the American Mathematical Society* 28: 669–709.

Balogh, J., R. Morris, and W. Samotij. 2018. "The Method of Hypergraph Containers." In *Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018*, Invited Lectures, vol. IV, 3059–3092. World Scientific Publishing.

Borgs, C., J. T. Chayes, H. Cohn, and Y. Zhao. 2019. "An L^p Theory of Sparse Graph Convergence I: Limits, Sparse Random Graph Models, and Power Law Distributions." *Transactions of the American Mathematical Society* 372: 3019–3062.

Chatterjee, S., and A. Dembo. 2016. "Nonlinear Large Deviations." *Advances in Mathematics* 299: 396–450.

Chatterjee, S., and S. R. S. Varadhan. 2011. "The Large Deviation Principle for the Erdős-Rényi Random Graph." *European Journal of Combinatorics* 32: 1000–1017.

Conlon, D., and W. T. Gowers. 2016. "Combinatorial Theorems in Sparse Random Sets." *Annals of Mathematics* 184, no. 2: 367–454.

Cook, N., and A. Dembo. 2020. "Large Deviations of Subgraph Counts for Sparse Erdős-Rényi Graphs." *Advances in Mathematics* 373: 107289.

Cook, N. A., A. Dembo, and H. T. Pham. 2021. "Regularity Method and Large Deviation Principles for the Erdős–Rényi Hypergraph".

Eldan, R. 2018. "Gaussian-Width Gradient Complexity, Reverse Log-Sobolev Inequalities and Nonlinear Large Deviations." *Geometric and Functional Analysis* 28: 1548–1596.

Frieze, A., and R. Kannan. 1999. "Quick Approximation to Matrices and Applications." *Combinatorica* 19: 175–220.

Janson, S., and L. Warnke. 2016. "The Lower Tail: Poisson Approximation Revisited." *Random Structures & Algorithms* 48: 219–246.

Kenyon, R., C. Radin, K. Ren, and L. Sadun. 2018. "Bipodal Structure in Oversaturated Random Graphs." *International Mathematics Research Notices* 2018, no. 4: 1009–1044.

Kozma, G., and W. Samotij. 2023. "Lower Tails via Relative Entropy." *Annals of Probability* 51: 665–698.

Lovász, L. 2012. "Large Networks and Graph Limits." In *American Mathematical Society Colloquium Publications*, vol. 60. American Mathematical Society.

Neeman, J., C. Radin, and L. Sadun. 2023. "Typical Large Graphs With Given Edge and Triangle Densities." *Probability Theory and Related Fields* 186: 1167–1223.

Radin, C. 2018. "Phases in Large Combinatorial Systems." *Annales de l'Institut Henri Poincaré D* 5: 287–308.

Saxton, D., and A. Thomason. 2015. "Hypergraph Containers." *Inventiones Mathematicae* 201: 925–992.

Schacht, M. 2016. "Extremal Results for Random Discrete Structures." *Annals of Mathematics* 184, no. 2: 333–365.

Zhao, Y. 2017. "On the Lower Tail Variational Problem for Random Graphs." *Combinatorics, Probability and Computing* 26: 301–320.

Zhao, Y. 2023. *Graph Theory and Additive Combinatorics—Exploring Structure and Randomness*. Cambridge University Press.