Large Deviations for Erdős-Rényi Graphs and Hypergraphs

Pengda Liu

Department of Mathematics

Stanford University

Advisors

Dr. Nicholas Cook and Dr. Amir Dembo

In partial fulfillment of the honors requirements for the degree of

Bachelor of Science in Mathematics

May 2020

Abstract

The upper tail problem for subgraph counts has attracted a great deal of interest over the years. This problem studies the asymptotic behavior of the log probability of the event that a subgraph H appears a constant factor more than its expectation in the Erdős-Rényi graph $G_{n,p}$. The problem has served as a testbed for developments in the emerging field of *nonlinear large deviations*, bringing together many areas of mathematics such as graph theory, probability theory, functional analysis, statistical physics and topology. In its development we see ideas from classical large deviations theory as well as the introduction of many new tools such as Szemerédi's regularity lemma, graphon theory and high dimensional geometry. In this thesis we give an expository account of the recent progress made in this area, and also give new results for random hypergraphs. Specifically, we develop an *induced hypergraph method* to prove a large deviation principle for counts of $K_{k+1}^{(k)}$ in the k-uniform Erdős-Rényi hypergraph by reducing the upper tail problem to a combinatorial optimization problem.

Acknowledgements

First and foremost, I am very grateful to my advisors, Dr. Nicholas Cook and Dr. Amir Dembo for their help and guidance. I am most indebted to Nick, who for the past year year has been making him available for our weekly meetings despite many other commitments. I feel very fortunate to have learnt a great deal from him. I learnt to start with simple cases both for understanding others' work and approaching new problems; I learnt to keep the big picture in mind instead of getting lost in the technical details. His positive outlook and calm passion for math left a deep impression on me. I thank my parents who constantly support me from far away for the last four years. I could not have attended Stanford and spent wonderful time here without their care and guidance on the way. In a sense everything starts with their confidence in me so I would like to dedicate this thesis to them even though they probably would not be able to understand it. Lastly, I thank Anita for her accompany and making life under COVID-19 and all sorts of pressure sweet. Time with her is 99% of laughter, 1% of debate and 100% valuable.

For my parents.

Contents

Chapter 1

Introduction

1.1 The goal of large deviations theory

Large deviation theory (shortened as LDT) studies the probability of certain rare events and more specifically, the asymptotic behavior of the log probability. A common scenario would be the upper tail probability of random variables. The following is an illustrative example taken from [9].

Example 1.1.1. *Let $X_1, X_2, ...,$ be i.i.d Bernoulli($\frac{1}{2}$) random variables and let $\hat{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$, then we have*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big(\hat{X}_n \geq (1+\frac{1}{3})\mathbb{E}[\hat{X}_n]\big) = -\log(2^{5/3}/3)).$$

This is saying that the probability of the upper tail event that $\hat{X}_n$ exceeds its expectation by a factor of at least $\frac{4}{3}$ is $e^{-0.0567...(1+o(1))n}$,which is decaying very fast. It will appear again in a more general form as Example 1.2.1 and be discussed later. The general definition for large deviation principle on topological space is given in the next section. The following class of problems is of great interest. Let $f : [0,1]^n \to \mathbb{R}$ be a smooth function and $X = (X_1, ..., X_n)$ be a vector of *i.i.d* Bernoulli(p) random variables. For $x = (x_1, ..., x_n) \in [0,1]^n$ define

$$I_p(x) = \sum_{i=1}^n I_p(x_i) \text{ where } I_p(x_i) = x_i \log \frac{x_i}{p} + (1-x) \log \frac{1-x_i}{1-p}. \tag{1.1}$$

It turns out in many situations we have that the upper tail probability can be reduced to the solution of a variational problem

$$\log \mathbb{P}(f(X) \geq (1+\delta)\mathbb{E}[X]) = -(1+o(1)) \inf_{x\in[0,1]^n} \{I_p(x) | f(x) \geq (1+\delta)\mathbb{E}[X]\} \qquad (1.2)$$

Note that the right hand side of Example 1.1.1 is equal to $I_{\frac{1}{2}}(\frac{2}{3})$. In the next section on classical large deviations theory, it will be illustrated that Equation (1.2) holds in great generality for linear function f such as $f(x) = \frac{x_1+...+x_n}{n}$. We will see later that Equation (1.2) also holds for some nonlinear functions f, though this can be challenging to prove.

1.2 Classical large deviations theory

Let's start with a motivating example slightly more general than Example 1.1.1 with $\frac{1}{3}$ replaced by an arbitrary constant $\delta > 0$.

Example 1.2.1. *Let $X_1, X_2, ...,$ be i.i.d Bernoulli(p) random variables and let $\hat{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$, then we have*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big(\hat{X}_n \geq (1+\delta)\mathbb{E}[\hat{X}_n]\big) = -\inf_{x\geq(1+\delta)\mathbb{E}[\hat{X}_n]} I_p(x) = -I_p((1+\delta)p).$$

The proof requires proving the upper bound and the lower bound. We first introduce the proof of the upper bound here which is a standard moment argument. We will prove the lower bound later in section 1.2.1.2 when we introduce a technique called the *tilting argument*.

Let $S_n = \sum_{i=1}^n X_i$ and $t = (1+\delta)\mathbb{E}[\hat{X}_n]$. Define the *logarithmic moment generating function*

$$\Lambda(\lambda) := \log \mathbb{E}[\exp(\lambda X_1)] = \log((1-p) + p\exp(\lambda)).$$

Note that for any $\lambda > 0$, we have

$$\begin{aligned} \mathbb{P}\big(\hat{X}_n \geq t\big) &= \mathbb{P}\big(S_n \geq nt\big) = \mathbb{P}\big(\exp(\lambda S_n) \geq \exp(\lambda nt)\big) \\ &\leq \frac{\mathbb{E}[\exp(\lambda S_n)]}{\exp(\lambda nt)} \qquad \text{(Markov's Inequality)} \\ &= \frac{\prod_{i=1}^n \mathbb{E}[\exp(\lambda X_i)]}{\exp(\lambda nt)} = \frac{\exp(\Lambda(\lambda)n)}{\exp(\lambda nt)} = \exp(n\Lambda(\lambda) - \lambda nt). \end{aligned}$$

By a direct calculation we have $\sup_{\lambda\geq 0}\{\lambda t-\Lambda(\lambda)\}=\sup_{\lambda\in\mathbb{R}}\{\lambda t-\Lambda(\lambda)\}$, so

$$\lim_{n\to\infty}\sup\frac{1}{n}\log\mathbb{P}(\hat{X}_n\geq t)\leq-\sup_{\lambda\in\mathbb{R}}\{\lambda t-\Lambda(\lambda)\}. \tag{1.3}$$

Solving the right side of (1.3) gives the upper bound for Example 1.2.1. This example dates back as early as Laplace's method for estimating integrals of the form

$$\int_a^b e^{Mf(x)}dx, \quad f \text{ twice differentiable and } M \text{ a large constant.}$$

Originating from this, the "right" level of generality for the LDT is due to Varadhan([32]) who proposed that similar formulation could be made about general topological measure space. We now formally introduce the concepts in LDT. Let $\mathcal{X}$ be a topological space with Borel σ-field $\mathcal{B}$.

Definition 1.2.2. *A rate function I is a lower semi-continuous function $I:\mathcal{X}\to[0,\infty]$. The effective domain is $D_I:=\{x:I(x)<\infty\}$.*

An example of a rate function is the right hand side of Example 1.2.1. The large deviation principle describes the asymptotic behavior of a sequence of measures $\{\mu_\epsilon\}$ as $\epsilon\to 0$.

Definition 1.2.3. *$\{\mu_\epsilon\}$ satisfies the large deviation principle with a rate function I if for all $\Gamma\in\mathcal{B}$,*

$$-\inf_{x\in\Gamma^o}I(x)\leq\lim_{\epsilon\to 0}\inf\epsilon\log\mu_\epsilon(\Gamma)\leq\lim_{\epsilon\to 0}\sup\epsilon\log\mu_\epsilon(\Gamma)\leq-\inf_{x\in\bar{\Gamma}}I(x). \tag{1.4}$$

The above definition allows a continuous family of measures in order to provide full generality and be applied to situations of continuous random processes like Brownian motions, see [15] for more account of this. In this thesis however we mainly care about a discrete sequence of measures $\{\mu_n\}$ and the behavior of $\lim a_n\log\mu_n$ for some sequence $\{a_n\}$ and we usually take Γ to be the upper tail event. In Example 1.2.1 we have $a_n=\frac{1}{n}$ and $\mu_n=\mathbb{P}(\hat{X}_n\geq(1+\delta)\mathbb{E}[\hat{X}_n])$.
Example 1.2.1 can be generalized to $i.i.d$ random variables in $\mathbb{R}^d$. We first make some formal definitions of the notions that appeared in Example 1.2.1.

Definition 1.2.4. *For a random variable X distributed according to μ on $\mathbb{R}^d$, the logarithmic moment generating function is*

$$\Lambda_\mu(\lambda) := \log \mathbb{E}[e^{\langle \lambda, X\rangle}] = \log \int_{\mathbb{R}^d} e^{\langle \lambda, x\rangle} \mu(dx).$$

We use the notation Λ when the random variable is clear under context.

The optimization problem on the right hand side of Equation (1.3) giving the rate function is an example of the following transformation.

Definition 1.2.5. *For a function $f : \mathbb{R}^d \to \mathbb{R}$, its Fenchel-Legendre transform is*

$$f^*(x) := \sup_{\lambda \in \mathbb{R}^d} \{\langle \lambda, x\rangle - f(\lambda)\}.$$

Let $X_1, X_2, ...,$ be a sequence of $i.i.d$ random variables in $\mathbb{R}^d$ with law μ and logarithmic moment generating function Λ. Let μ_n be the law of $\hat{X}_n = \frac{X_1 + ... X_n}{n}$. The following theorem gives a general large deviation principle for $\{\mu_n\}$.

Theorem 1.2.6 (Cramér, Theorem 2.2.30 in [15])**.** *Assume that $D_\Lambda = \mathbb{R}^d$, that is $\Lambda(\lambda) < \infty$ for all $\lambda \in \mathbb{R}^d$. Then $\{\mu_n\}$ satisfies the large deviation principle with rate function Λ^*.*

1.2.1 Abstract LDT in topological vector space

In this section, we introduce some large deviation results that hold abstractly, relaxing the $i.i.d$ and $\mathbb{R}^d$ assumptions. Let $\mathcal{X}$ be a Hausdorff topological vector space and $\mathcal{X}^*$ be its dual space, namely, the space of all continuous linear functionals on $\mathcal{X}$. Let $\{X_\epsilon\}$ be a family of random variables taking values in $\mathcal{X}$ with law μ_ϵ. Analogous to the $\mathbb{R}^d$ case, the *logarithmic moment generating function* $\Lambda_{\mu_\epsilon}(\lambda) : \mathcal{X}^* \to [-\infty, \infty]$ is

$$\Lambda_{\mu_\epsilon} = \log \mathbb{E}[e^{\langle \lambda, X_\epsilon\rangle}] = \log \int_{\mathcal{X}} e^{\lambda(x)} \mu_\epsilon(dx), \qquad \lambda \in \mathcal{X}^*.$$

where for $x \in \mathcal{X}$ and $\lambda \in \mathcal{X}^*$, $\langle \lambda, x\rangle$ denotes the value of $\lambda(x) \in \mathbb{R}$.
Let

$$\bar{\Lambda}(\lambda) = \limsup_{\epsilon \to 0} \epsilon \Lambda_{\mu_\epsilon}(\frac{\lambda}{\epsilon}) \tag{1.5}$$

and use the notation $\Lambda(\lambda)$ whenever the limit exists, which is the case for the examples in the previous subsections for $i.i.d$ random variables in $\mathcal{X} = \mathbb{R}^d$. where we have that

$\frac{1}{n}\Lambda_{\mu_n}(n\lambda)$ converges to Λ_μ. Definition 1.2.5 generalizes to this setting as follows:

Definition 1.2.7. *For a function $f : \mathcal{X}^* \to [-\infty, \infty]$, its Fenchel-Legendre transform is*

$$f^*(x) := \sup_{\lambda \in \mathcal{X}^*} \{\langle \lambda, x\rangle - f(\lambda)\}, \qquad x \in \mathcal{X}.$$

1.2.1.1 Upper bound with covering

We have the following general upper bound. Recall that $\bar{\Lambda}^*$ is the Legendre transform of $\bar{\Lambda}$ defined in (1.5).

Theorem 1.2.8 (Theorem 4.5.3 in [15]). *For any compact set $\Gamma \in \mathcal{X}$,*

$$\limsup_{\epsilon \to 0} \epsilon \log \mu_\epsilon(\Gamma) \leq - \inf_{x \in \Gamma} \bar{\Lambda}^*(x).$$

The rough idea of the proof is as follows. We first prove this for certain small open neighborhood $\mathcal{B}$ using a similar moment method as in the proof of the upper bound of Example 1.2.1. Then by compactness we can find a finite cover $\cup \mathcal{B}_i$ and the result follows by a union bound. We can replace the compact set with a closed set if we assume the following condition on $\{\mu_\epsilon\}$ that roughly says they are concentrating on compact sets.

Definition 1.2.9. *A family of measure $\{\mu_\epsilon\}$ on $\mathcal{X}$ is said to be exponentially tight if for any $\alpha > 0$, there is a compact subset $\mathcal{K}_\alpha \subset \mathcal{X}$ such that*

$$\limsup_{\epsilon \to 0} \epsilon \log \mu_\epsilon(\mathcal{X} \setminus \mathcal{K}_\alpha) < -\alpha.$$

Corollary 1.2.10. *Assume that $\{\mu_\epsilon\}$ is exponentially tight. Then for any closed set $F \subset \mathcal{X}$,*

$$\limsup_{\epsilon \to 0} \epsilon \log \mu_\epsilon(F) \leq - \inf_{x \in F} \bar{\Lambda}^*(x).$$

1.2.1.2 Lower bound with tilting argument

A common technique for proving the lower bound in a large deviation principle is the following tilting argument. For $\Gamma \subset \mathcal{X}$ and a measure $\mathbb{P}$, to bound $\mathbb{P}(\Gamma)$, we first find a dominating event $\mathcal{B} \subset \Gamma$ and construct a measure $\mathbb{Q}$ under which $\mathcal{B}$ is typical with

$\mathbb{Q}(\mathcal{B}) \approx 1$. Then we have

$$\mathbb{P}(\mathcal{B}) = \int_{\mathcal{B}} d\mathbb{P} = \int_{\mathcal{B}} \frac{d\mathbb{P}}{d\mathbb{Q}} d\mathbb{Q} = \mathbb{Q}(\mathcal{B}) \frac{1}{\mathbb{Q}(B)} \int_{\mathcal{B}} \frac{d\mathbb{P}}{d\mathbb{Q}} d\mathbb{Q}$$

where $\frac{d\mathbb{P}}{d\mathbb{Q}}$ is the Radon-Nikodym derivative. Thus by Jensen's Inequality

$$\log \mathbb{P}(\Gamma) \geq \log \mathbb{P}(\mathcal{B}) = \log \mathbb{Q}(\mathcal{B}) - \log \Big(\frac{1}{\mathbb{Q}(B)} \int_{\mathcal{B}} \frac{d\mathbb{P}}{d\mathbb{Q}} d\mathbb{Q} \Big)$$

$$\geq \log \mathbb{Q}(\mathcal{B}) - \frac{1}{\mathbb{Q}(B)} \int_{\mathcal{B}} \log \frac{d\mathbb{P}}{d\mathbb{Q}} d\mathbb{Q} \approx - \int_{\mathcal{B}} \log \frac{d\mathbb{P}}{d\mathbb{Q}} d\mathbb{Q}.$$

The challenge is to choose the right measure $\mathbb{Q}$ so that the integral will give a sharp lower bound. As an example, we prove the lower bound of Example 1.2.1.

Recall that $X_1, ..., X_n$ are $i.i.d$ Bernoulli(p) random variables with law $\mathbb{P}$ and average $\hat{X}_n$. Let $\delta > 0$ and set $t = (1+\delta)\mathbb{E}[\hat{X}_n]$. Let $\epsilon, \lambda > 0$ to be decided. Define another measure

$$d\mathbb{Q}(X) = e^{-\Lambda(\lambda)+\lambda X} d\mathbb{P}(X) \text{ and } d\mathbb{Q}_n = d\mathbb{Q} \otimes ... \otimes d\mathbb{Q}.$$

Let $d\mathbb{P}_n = d\mathbb{P} \otimes ... \otimes d\mathbb{P}$, we have

$$d\mathbb{Q}_n = e^{-n\Lambda(\lambda)+\lambda(X_1+...+X_n)} d\mathbb{P}_n.$$

Let $\mathcal{B}_{n,\epsilon} = \{t \leq \hat{X}_n < t + \epsilon\}$. Since

$$\mathbb{E}_{\mathbb{Q}_n}[\hat{X}_n] = \Lambda'(\lambda) = e^{-\Lambda(\lambda)} \mathbb{E}[X e^{\lambda X}],$$

and hence $\Lambda'(0) = \mathbb{E}[\hat{X}_n] < t, \Lambda'(\infty) = 1 > t$, assuming $p < \frac{1}{1+\delta}$ when we let $\epsilon \to 0$ by intermediate value theorem we can find always λ so that $\mathbb{E}_{\mathbb{Q}_n}[\hat{X}_n] = t + \frac{\epsilon}{2}$. Then by the law of large numbers $\mathbb{Q}_n$ is concentrated on $\mathcal{B}_{n,\epsilon}$, meaning that $\lim_{n\to\infty} \mathbb{Q}_n(\mathcal{B}_{n,\epsilon}) = 1$. Let $\mathcal{B} = \mathcal{B}_{n,\epsilon}$ in the tilting argument we described above,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_n(\hat{X}_n \geq t) \geq \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_n(\mathcal{B}_{n,\epsilon}) = -\frac{1}{n} \int_{\mathcal{B}_{n,\epsilon}} \big(-n\Lambda(\lambda)+\lambda(X_1+...+X_n)\big) d\mathbb{Q}_n$$

$$\geq -(\lambda(t+\epsilon) - \Lambda(\lambda)) \geq - \sup_{\lambda'>0} \{\lambda'(t+\epsilon) - \Lambda(\lambda')\}.$$

By letting $\epsilon \to 0$,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_n(\hat{X}_n \geq t) \geq -\lim_{\epsilon\to 0} \sup_{\lambda'>0}\{\lambda'(t+\epsilon) - \Lambda(\lambda')\} = -\Lambda^*(t),$$

thus establishing the lower bound for Example 1.2.1. We will see another application of this technique in section 2.4.2. We now state the following theorem on the lower bound in abstract LDT whose proof roughly follows what we described above.

Definition 1.2.11. *A point $x \in \mathcal{X}$ is called an exposed point of $\bar{\Lambda}^*$ if there exists an exposing hyperplane $\lambda \in \mathcal{X}^*$ such that*

$$\langle \lambda, x\rangle - \bar{\Lambda}^*(x) > \langle \lambda, z\rangle - \bar{\Lambda}^*(z), \qquad \forall z \neq x.$$

Theorem 1.2.12 (Baldi, Theorem 4.5.20 in [15])**.** *Let $\{\mu_\epsilon\}$ be exponentially tight measures and $\mathcal{F}$ be the set of exposed points of $\bar{\Lambda}^*$ with an exposing hyperplane λ for which*

$$\Lambda(\lambda) = \lim_{\epsilon\to 0} \epsilon \Lambda_{\mu_\epsilon}(\frac{\lambda}{\epsilon}) \text{ exists and } \bar{\Lambda}(\gamma\lambda) < \infty \text{ for some } \gamma > 1.$$

Then for every open set $G \subset \mathcal{X}$,

$$\liminf_{\epsilon\to 0} \log \mu_\epsilon(G) \geq -\inf_{x\in G\cap\mathcal{F}} \bar{\Lambda}^*(x).$$

If we further assume that for every open set G,

$$\inf_{x\in G\cap\mathcal{F}} \bar{\Lambda}^*(x) = \inf_{x\in G} \bar{\Lambda}^*(x),$$

combining with Corollary 1.2.10 we get a large deviation principle for $\{\mu_\epsilon\}$ on $\mathcal{X}$.
Linear large deviation theory relies on the fact that the logarithmic moment generating function $\Lambda(\lambda)$ is easy to evaluate and work with, as we see in the proof of Example 1.2.1. In section 1.4, we introduce the case when the function is subgraph counts and difficulties arise.

1.3 Connection of LDT to statistical physics

In statistical physics, a common problem is to estimate the partition function of a system

$$Z = \sum_x e^{-\beta E(x)},$$

where the sum is over all microstates and $E(x)$ is the energy in each state. Recall that β is the thermodynamic beta defined as $\beta = \frac{1}{k_{\mathcal{B}}T}$ where $k_{\mathcal{B}}$ is the Boltzmann constant and T is the temperature. The following Varadhan's Integral Lemma shows that a large deviation principle implies the convergence of certain partition function. As in the previous set up, let $\{X_\epsilon\}$ be a family of random variables taking values in the topological space $\mathcal{X}$ with associated measures $\{\mu_\epsilon\}$.

Theorem 1.3.1 (Theorem 4.3.1 in [15], Varadhan)**.** *Suppose that $\{\mu_\epsilon\}$ satisfies a large deviation principle with rate function $I : \mathcal{X} \to [0, \infty]$, and let $\phi : \mathcal{X} \to \mathbb{R}$ be any continuous function. Assume further either the tail condition*

$$\lim_{M\to\infty} \limsup_{\epsilon\to 0} \epsilon \log \mathbb{E}[e^{\phi(X_\epsilon)/\epsilon} \mathbf{1}_{\{\phi(X_\epsilon)\geq M\}}] = -\infty,$$

or the following moment condition for some $\gamma > 1$,

$$\limsup_{\epsilon\to 0} \epsilon \log \mathbb{E}[e^{\gamma\phi(X_\epsilon)/\epsilon}] < \infty.$$

Then

$$\lim_{\epsilon\to 0} \epsilon \log \mathbb{E}[e^{\phi(X_\epsilon)/\epsilon}] = \sup_{x\in X}\{\phi(x) - I(x)\}.$$

The other direction (the convergence of certain partition function implies a large deviation principle) is given by the following Bryc's Inverse Varadhan Lemma. Let $C_b(\mathcal{X})$ be set of bounded and continuous real-valued functions on $\mathcal{X}$. For $f \in C_b(\mathcal{X})$, define

$$\Lambda_f := \lim_{\epsilon\to 0} \epsilon \log \int_{\mathcal{X}} e^{f(x)/\epsilon} \mu_\epsilon(dx), \tag{1.6}$$

provided the limit exists.

Theorem 1.3.2 (Theorem 4.4.2 in [15], Bryc)**.** *Suppose that $\{\mu_\epsilon\}$ is exponentially tight and that the limit Λ_f in* (1.6) *exists for every $f \in C_b(\mathcal{X})$. Then $\{\mu_\epsilon\}$ satisfies the large* *deviation principle with rate function*

$$I(x) = \sup_{f \in C_b(\mathcal{X})} \{f(x) - \Lambda_f\}.$$

Furthermore, for every $f \in C_b(\mathcal{X})$,

$$\Lambda_f = \sup_{x \in \mathcal{X}} \{f(x) - I(x)\}.$$

If $\mathcal{X}$ is a topological vector space and $f \in \mathcal{X}^*$, then Λ_f is just the logarithmic moment generating function defined in (1.5). Next, we illustrate further connections with the example of Gibbs measure.

1.3.1 Gibbs measure on the Hamming cube

Given a Hamiltonian $f : \{0,1\}^d \to \mathbb{R}$, we can associate it with a Gibbs measure μ on $\{0,1\}^d$ with density $Z^{-1} e^{f(\cdot)}$. Then

$$Z = \sum_{x \in \{0,1\}^d} e^{f(x)}.$$

Determining the large deviations of a function h of a random Bernoulli(p) on $\{0,1\}^d$ can be formulated as estimating the partition function Z. That is, approximating

$$\log \mathbb{P}(h(x) \geq (1+\delta)\mathbb{E}[h(x)])$$

corresponds to estimating $\log Z$ for

$$f_h(x) := g(h(x)) + d \log(1-p) + \sum_{i=1}^{d} x_i \log \frac{p}{1-p},$$

where $g(s) \equiv 0$ for $s \geq (1+\delta)\mathbb{E}[h(x)]$ and $g(s) \equiv -\infty$ for $s < (1+\delta)\mathbb{E}[h(x)]$. By the Gibbs variational principle, we have

$$\log Z = \sup_{\mu \in M_1(\{0,1\}^d)} \{\sum_x f(x)\mu(x) - \sum_x \mu(x) \log \mu(x)\},$$

where $M_1(\{0,1\}^d)$ is the set of all probability measures on the cube.
Since the dimension of $M_1(\{0,1\}^d)$ is exponential in d, it is common practice in physics to use the *naive mean field* approximation, which is to restrict μ to product measures. Note that when the Hamiltonian f is separable, meaning that it takes the form

$$f(x) = f_1(x_1) + ... + f_d(x_d)$$

(such as when f is linear), the naive mean field approximation is an exact identity.

1.4 The upper tail problem in subgraph counts

Throughout this thesis we refer only to simple graphs where there is no self loop or multi-edge. Let $G_{n,p}$ be the Erdős-Rényi graph model on vertices $[n]$ where the edge connections are *i.i.d* Bernoulli(p) random variables. For a fixed graph H, an interesting quantity is $N(H, G)$, the number of subgraphs H present in G where G is sampled from $G_{n,p}$. The simplest non-trivial case is the number of triangles. Let x_{ij} be the indicator random variable for edge connection between i and j of G, we have

$$N(K_3, G) = \sum_{1 \le i < j < k \le n} x_{ij} x_{jk} x_{ik}. \tag{1.7}$$

Let Y_H be the random variable $N(H, G)$, the "*infamous upper tail problem*" [23] asks to estimate the upper tail probability

$$R(H, n, p, \delta) := \log \mathbb{P}(Y_H \ge (1 + \delta)\mathbb{E}[Y_H]). \tag{1.8}$$

This is a classical problem and has been extensively studied (see [8, 13, 14, 22–25, 33] and [4] and its references). The goal is to solve this for general H, however, even for $H = K_3$ where it is just a polynomial of degree 3 seems to be highly non-trivial and beyond the capabilities of classical large deviations theory introduced in section 1.2. Let's give a brief account of the older works that use concentration machinery instead. [33] and [25] showed that

$$n^2 p^2 \lesssim R(K_3, n, p, \delta) \lesssim n^2 p^2 \log(1/p).$$

The correct order of the rate function $R(K_3, n, p, \delta) \asymp n^2p^2 \log(1/p)$ was independently established in [8, 14]. However, the methods of these works did not allow determining the exact asymptotics of the rate function or extending to general graphs.

We now introduce an alternative and sometimes easier to deal with metric for subgraph count that we will use a bit more frequently. Let $\mathrm{Hom}(H, G)$ be the set of functions $f : V(H) \to V(G)$ so that $(f(i), f(j)) \in E(G)$ for every $(i, j) \in E(H)$. The *homomorphism density function* is defined to be

$$t(H, G) := \frac{|\mathrm{Hom}(H, G)|}{|V(G)|^{|V(H)|}}. \tag{1.9}$$

There are standard relations between $N(H, G)$ and $|\mathrm{Hom}(H, G)|$ (see [28]) and their multiplicative difference is negligible in the setting we work with. Let X_H be the random variable $t(H, G_{n,p})$. By linearity of expectation, we have $\mathbb{E}[X_H] \sim p^{|E(H)|}$. Similar to (1.8), we are interested in estimating the following log upper tail probability for X_H,

$$\zeta(H, n, p, \delta) := \log \mathbb{P}(X_H \geq (1 + \delta)p^{|E(H)|}) \text{ for fixed } \delta > 0. \tag{1.10}$$

Let $\mathcal{G}_n$ be the set of edge-weighted undirected graphs on n vertices, that is, if $A(G)$ is the adjacency matrix of a graph G then

$$\mathcal{G}_n = \{G : A(G)_{ij} \in [0, 1], A(G)_{ij} = A(G)_{ji}, A(G)_{ii} = 0 \text{ for all } 1 \leq i, j \leq n\} \tag{1.11}$$

We can extend the homomorphism counting function in (1.9) to $G \in \mathcal{G}_n$ as

$$t(H, G) := n^{-|V(H)|} \sum_{\iota:[k]\to[n]} \prod_{(x,y)\in E(H)} A(G)_{\iota(x)\iota(y)} \tag{1.12}$$

When we fix a graph H, we use the notation $t_H(\cdot) := t(H, \cdot)$. The function t_H is a nonlinear polynomial of degree e_H but as we mentioned earlier, even the simplest case of $H = K_3$ where it is just degree 3 has great difficulties. In the next section, we introduce some significant progress that were made fairly recently, taking a different route from classical concentration inequalities.

1.5 Outline for recent progress

In the last section, we see that it seems that one could not solve the subgraph count problem in the framework of large deviation theory due to its non-linearity. However, since the breakthrough made by [11], a series of works [2,3,10,12,16,21] has introduced lots of new tools and ideas to make this possible. First we introduce the general route taken by these works. There are two major steps. The first step is similar to the results of classical large deviations theory we introduced in section 1.2 and attempts to reduce the upper tail $\zeta(H,n,p,\delta)$ in (1.10) to a variational problem. By section 1.1, the natural variational problem to consider is

$$\phi(H,n,p,\delta) := \inf_{G\in\mathcal{G}_n}\{I_p(G) : t_H(G) \geq (1+\delta)p^{|E(H)|}\} \tag{1.13}$$

recall that I_p defined in (1.1) and $I_p(G) := \sum_{1\leq i<j\leq n} I_p(A(G)_{ij})$. So the goal of the first step is to prove a result of the form

$$\zeta(H,n,p,\delta) = -(1+o(1))\phi(H,n,p,\delta+o(1)). \tag{1.14}$$

We refer to this step as *variational reduction*. The second step is to solve the asymptotic form of (1.13) explicitly, which we refer to as *variational solution*. Next we give a brief introduction to these works mostly using the example of $H = K_3$ and we will describe their results in greater detail in later chapters.

1.5.1 First step: variational reduction

The problem is divided into two regimes. One is the dense regime where p is fixed and another is the sparse regime where $p = p(n)$ decays to 0 and $n^{-\alpha} \ll p \ll 1$ for some constant $\alpha > 0$. The dense regime is solved by [11] with an amazing use of the *graphon theory* and Szemerédi's *regularity lemma*. Their main idea is to cover the "space of graphs" with sets constructed from regularity lemma on which we can apply the results from classical large deviations theory in section 1.2 and then apply a union bound. However, due to the bad quantitative dependency of the regularity lemma, this method does not apply to the sparse regime and two lines of ideas are introduced.

The first line is the work of [10,16]. A non-linear large deviation framework based on the *low gradient complexity* condition was established in [10] and they were able to apply their technique to prove (1.14) for $H = K_3$ in the range $p \gg n^{-1/42}(\log n)^{11/14}$. Their work was further developed by [16] which proved (1.14) in the range $p \gg n^{-1/18} \log n$. The main idea behind the work of [10, 16] is inspired by statistical physics introduced in section 1.3, showing that the naive mean field approximation holds not only when the Hamiltonian is separable, but also when the gradient ∇t_H can be approximated using a net.

The second line is the work of [12, 21]. Their work can be seen as a more quantitative continuation of [11]. Instead of approximating the gradient, their proof is based on efficient coverings of the space. Using spectral method, [12] covered the cube $\{0,1\}^{\binom{n}{2}}$ with convex bodies on which t_H is nearly constant. After obtaining a sharp estimate of each convex body, they take the union to get an estimate of the upper tail set. They showed (1.14) for $H = K_3$ in the range $p \gg n^{-1/3}$.

More recently, [21] introduced the following combinatorial optimization problem

$$\Phi(H,n,p,\delta) = \min\{e'_G \log(1/p) : G' \subset K_n \text{ and } \mathbb{E}_{G'}[N(H,G)] \geq (1+\delta)\mathbb{E}[N(H,G)]\}, \tag{1.15}$$

where $\mathbb{E}_{G'}[N(H,G)] = \mathbb{E}[N(H,G)|G' \subset G]$. As we will see later, Φ and ϕ behave the same asymptotically. The work of [21] established an *entropic stability* framework based on covering the space with near-optimizers of the above optimization problem. Their method is more combinatorial and they worked with subgraph count function $N(H,G)$ and R in (1.8) instead of $t(H,G)$ and ζ. They proved that for $p \gg n^{-1} \log n$ we have

$$R(K_3,n,p,\delta) = -(1+o(1))\Phi(K_3,n,p,\delta+o(1)), \tag{1.16}$$

also showing this is essentially the optimal range.

In later chapters we will mainly focus on the approach of the second line ([11, 12, 21]). For more detailed discussion of the first line([10, 16]), please refer to their papers and the survey [9] by Chatterjee.

1.5.2 Second step: variational solution

For a fixed p, there are no known explicit solution to the variational problem for general H except for under very special situations. For $H = K_3$ we do have an explicit solution from [11],

$$\phi(K_3, n, p, \delta) = \frac{1}{2}n^2 f((1+\delta)p^3)(1+o(1))$$

where

$$f(t) = \begin{cases} -\log 2 & \text{if } 0 < t < \frac{1}{48}, \\ I_p((6t)^{1/3}) & \text{if } \frac{1}{48} \le t \le \frac{1}{6}, \\ +\infty & \text{if } t > \frac{1}{6}. \end{cases}$$

For the sparse regime [29] showed that

$$\lim_{n\to\infty} \frac{\phi(K_3, n, p, \delta)}{n^2 p^2 \log(1/p)} = \begin{cases} \min\{\frac{\delta^{2/3}}{2}, \frac{\delta}{3}\} & \text{if } n^{-1/2} \ll p \ll 1, \\ \frac{\delta^{2/3}}{2} & \text{if } n^{-1} \ll p \ll n^{-1/2}. \end{cases} \tag{1.17}$$

To gain some insight on where this formula comes from, consider the graph $G \in \mathcal{G}_n$ starting with all weights set to p and modified by setting $A(G)_{ij} = 1$ for $i, j \le s$ where $s \sim \delta^{1/3} np$. Then G contains an extra of $\delta\mathbb{E}[X_{K_3}]$ triangles and we have $I_p(G) \sim \frac{\delta^{2/3}}{2} n^2 p^2$, thus giving an upper bound on ϕ. More detailed construction and analysis is given in Chapter 5. It was shown in [21] that this is the asymptotic solution to the combinatorial optimization problem Φ in (1.15) as well. Combining the two steps together we have the following result for the spare regime of K_3.

Theorem 1.5.1 (Theorem 1.7 in [21]). *For any $\delta > 0$, if $n^{-1}\log n \ll p \ll 1$, then*

$$\lim_{n\to\infty} \frac{\log \mathbb{P}(N(K_3, G) \ge (1+\delta)\mathbb{E}[N(K_3, G)])}{n^2 p^2 \log(1/p)} = \begin{cases} -\min\{\frac{\delta^{2/3}}{2}, \frac{\delta}{3}\} & \text{if } n^{-1/2} \ll p \ll 1, \\ -\frac{\delta^{2/3}}{2} & \text{if } n^{-1} \ll p \ll n^{-1/2}. \end{cases} \tag{1.18}$$

Since $\mathbb{E}[X_{K_3}] \asymp n^3 p^3$, this range is essentially optimal(ignoring log factors) since for p below this range the number of triangles is bounded and X_{K_3} behaves like a Poisson random variable(see [21] for more about this regime). Thus we completed the story of the upper tail probability of subgraph count of K_3 except for $np \to c$ for a constant c (see [21] for discussions of this edge case). It was later extended in [3] to solve the variational problem for general H; we defer discussion of this to Chapter 5.

1.6 Research contribution in hypergraphs

The natural next question is to solve the upper tail problem for hypergraphs. Hypergraphs are generalizations of graphs where edges are defined on possibly more than 2 vertices. We restrict our attention to r-uniform hypergraphs where each edge consists of a fixed number r of vertices. A graph is a 2-uniform hypergraph and we will extend some of the results in the previous section to a type of hypergraph cliques. We take the two step route similar to the graph case. There have been few works([26, 30]) done in this area and they mainly focus on the variational problem ϕ and less on the first step which is to reduce the upper tail to ϕ or Φ, which will be our focus. Specifically, we prove the following theorem.

Theorem 1.6.1. *For fixed $k \geq 3$ and $p = p(n)$ such that $n^{-1} \log n \ll p \ll 1$,*

$$R(K_{k+1}^{(k)}, n, p, \delta) = -(1 + o(1))\Phi(K_{k+1}^{(k)}, n, p, \delta + o(1)). \tag{1.19}$$

Note that the upper tail R in (1.8) and the combinatorial optimization problem Φ in (1.15) can be defined analogously for Erdős-Rényi hypergraphs $G_{n,p}^{(k)}$ where each k-element subset of $[n]$ is included as an edge independently with probability p. We will provide more details in Chapter 6.

1.7 Outline of the thesis

In Chapters 2 to 4 we introduce the aforementioned work in greater detail and sketch their proof. Specifically, in Chapter 2 we introduce the background story of graphon theory and regularity lemma and we show how these are used by [11] to complete the variational reduction step in the dense regime. In Chapters 3 and 4 we explain the works [12] and [21] respectively on the sparse regime. In Chapter 5 we explain the work done by [29] and [3] in solving the variational problem in the sparse regime. Finally, in the last Chapter, we introduce our work done in the hypergraph setting and prove Theorem 1.6.1.

1.8 Notations

In this section we introduce some notations that are used through out this thesis. Other notations will be introduced in their chapters as needed.

$[n]$	The set of integers $\{1, ..., n\}$.
$V(G)$	The set of vertices of the graph G.
$E(G)$	The set of edges of the graph G.
v_G	The number of vertices of G, that is, $\lvert V(G)\rvert$.
e_G	The number of edges of G, that is, $\lvert E(G)\rvert$.
K_r	The clique graph with r vertices.
C_l	The cycle graph with l vertices.
$K_k^{(r)}$	The r-uniform hypergraph clique with k vertices.
$G_{n,p}$	The Erdős-Rényi graph on vertices $[n]$ with edge probability p.
$\mathcal{G}_n$	Defined in (1.11),the set of edge weighted graphs on vertices $[n]$.
$G_{n,p}^{(r)}$	The Erdős-Rényi r-uniform hypergraph on vertices $[n]$ with edge probability p.
$N(H, G)$	The subgraph counting function (the number of copies of J in G).
$\mathrm{Hom}(H, G)$	The set of graph homomorphism from H to G.
$t(H, G)$	Defined in (1.9) and (1.12), the homomorphism density function.
R	Defined in (1.8), the upper tail for subgraph counting function $N(H, G)$.

ζ	Defined in (1.10), the upper tail for homomorphism density function $t(H, G)$.
ϕ	Defined in (1.13), the variational problem .
Φ	Defined in (1.15), the combinatorial optimization problem.
X_H	The random variable $t(H, G_{n,p})$.
Y_H	The random variable $N(H, G_{n,p})$.
$\mathbf{1}_A$	The indicator function of the set A.
$f = O(g), g = \Omega(f)$	There is a constant C so that $\|f\| \leq Cg$
$f \lesssim g, g \gtrsim f$	Both synonymous to the above.
$f = \Theta(g)$	$f = O(g)$ and $f = \Omega(g)$.
$f \asymp g$	Synonymous to the above.
$f = o(g)$	$f/g \to 0$ as $n \to \infty$.
$f \ll g, g \gg f$	Both synonymous to the above.
$f \sim g$	$f = (1 + o(1))g$.
Γ^o	The interior of the set Γ.
$\bar{\Gamma}$	The closure of the set Γ.

Chapter 2

Variational reduction in the dense regime

In this chapter we outline important ideas that were introduced in [11], particularly their use of graphon theory and regularity lemma. One of the main points is that one can use the graphon topology(in particular the regularity lemma) to find coverings of the space of graphs by certain balls on which the homomorphism density function t_H is almost constant. The probability of each ball can be estimated using results from classical large deviations theory.

2.1 Graphon theory and graph limits

The first difficulty in formulating a large deviation principle for random graphs is to find a common space to include all graphs of different number of vertices and to have a natural limit object for a series of graphs. This is resolved by a sequence of works([5–7, 27, 28]) that developed a beautiful and unifying theory of *graphon theory* and *graph limits*. First we notice that a graph G on $[n]$ can be represented by a function on the $n \times n$ grid in $[0,1]^2$ given by

$$f^G(x,y) = \begin{cases} 1 & \text{if } (\lceil nx \rceil, \lceil ny \rceil) \in E(G), \\ 0 & \text{otherwise.} \end{cases} \tag{2.1}$$

An example is $f^{K_{2,2}}$ in Figure 2.1.

Let $\mathcal{W}$ be the space of symmetric measurable functions $f : [0,1]^2 \to [0,1]$ so that $f(x,y) = f(y,x)$ then from (2.1) we can embed any graph into this space. Let H be

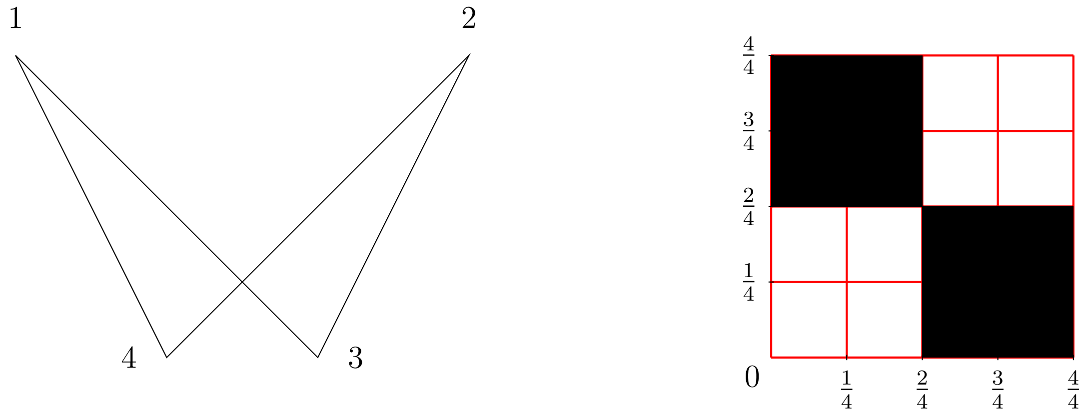

Figure 2.1 $K_{2,2}$ and $f^{K_{2,2}}$ with black region having value 1

a graph on k vertices, recall that in (1.9) we defined $t(H, G)$ to be the homomorphism density of H in G, we can also define this similarly for $f \in \mathcal{W}$ as

$$t(H, f) = \int_{[0,1]^k} \prod_{(i,j)\in E(H)} f(x_i, x_j)dx_1...dx_k. \tag{2.2}$$

We then have $t(H, f^G) = t(H, G)$ and with this association we can define the following graph limits.

Definition 2.1.1. *A sequence of graphs $\{G_n\}_{n=1}^{\infty}$ converges to $f \in \mathcal{W}$ if for every graph H we have $\lim_{n\to\infty} t(H, G_n) = t(H, f)$.*

This definition also provides a natural limit object for the sequence of Erdős–Rényi graph. With standard concentration inequalities, we have the following.

Proposition 2.1.2. *For a fixed $p \in [0, 1]$, let $f \in \mathcal{W}$ be the constant function such that $f(x, y) = p$ for all $(x, y) \in [0, 1]^2$. The sequence of Erdős–Rényi random graphs satisfy*

$$\lim_{n\to\infty} G_{n,p} = f \text{ with probability 1.}$$

We can define the following distance between "graphs" in $\mathcal{W}$.

Definition 2.1.3. *For $f, g \in \mathcal{W}$, the cut distance between them is*

$$d_\square(f, g) := \sup_{S,T\in\sigma([0,1])} \left| \int_{S\times T} [f(x, y) - g(x, y)]dxdy \right|.$$

Naturally we would like that two isomorphic graphs G_1 and G_2 should satisfy $d_\square(f^{G_1}, f^{G_2}) = 0$ but this is not currently achieved. Take the following two isomorphic graphs $G_1 = C_4$

and $G_2 = K_{2,2}$, then it's easy to see from Figure 2.2 that $d_\square(f^{G_1}, f^{G_2}) \geq \frac{1}{16}$ by taking $S \times T$ to be a single cell. This motivates the following equivalence relation on $\mathcal{W}$. Let Σ

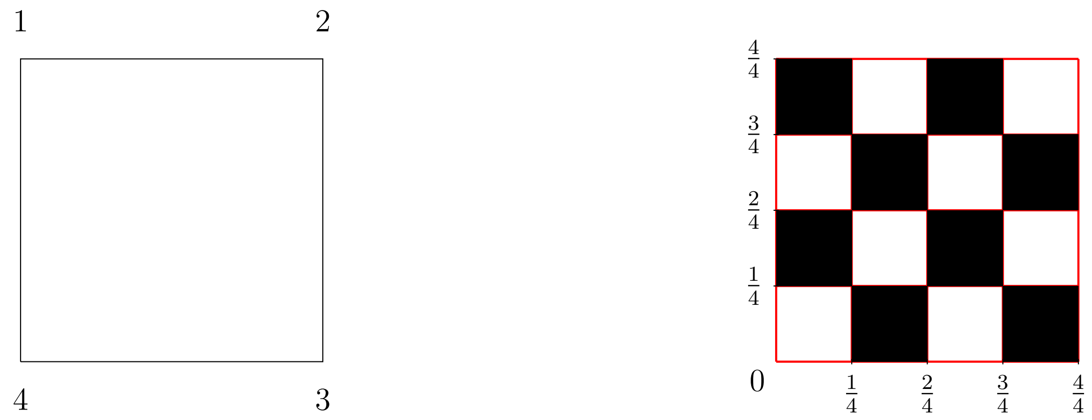

Figure 2.2 C_4 and f^{C_4} with black region having value 1

be the space of measure preserving bijections $\sigma : [0,1] \to [0,1]$. Note that the cut metric d is invariant under σ. Thus just like graph equivalence given by some bijection from $\{1, ..., n\} \to \{1, ..., n\}$, the equivalence relation on $\mathcal{W}$ is given by some mapping from σ.

Definition 2.1.4. *For $f \in \mathcal{W}$, let f_σ be defined as $f_\sigma(x,y) = f(\sigma x, \sigma y)$. Another $g \in \mathcal{W}$ satisfies $f \sim g$ if there exists $\sigma \in \Sigma$ such that $f_\sigma = g$.*

It's easy to see that this is an equivalent relation and $f^G \sim f^H$ for graphs $G \sim H$. Let $\tilde{f}$ be the closure of the orbit $\{f_\sigma\}$ of f and let τ be the map $f \to \tilde{f}$ we can now officially introduce the graphon space.

Definition 2.1.5. *The graphon space $\tilde{\mathcal{W}}$ is the quotient space $\mathcal{W} \setminus \tau$.*

Again by the invariance of the cut metric of $\mathcal{W}$ under σ, we can define the metric $\tilde{d}_\square$ on $\tilde{\mathcal{W}}$ as

$$\tilde{d}_\square(\tilde{f}, \tilde{g}) = \inf_\sigma d_\square(f, g_\sigma) = \inf_\sigma d_\square(f_\sigma, g) = \inf_{\sigma_1, \sigma_2} d_\square(f_{\sigma_1}, g_{\sigma_2}).$$

With the above formulation, for any graph G we can associate it a graphon by

$$G \to f^G \to \tilde{G} := \tilde{f}^G \in \tilde{\mathcal{W}}.$$

Then the Erdős–Rényi random graph $G_{n,p}$ would induce a measure $\mathbb{P}_{n,p}$ on $\mathcal{W}$ and $\tilde{\mathbb{P}}_{n,p}$ on $\tilde{\mathcal{W}}$. The function I_p defined in (1.1) could be extended to $\mathcal{W}$ as

$$I_p(f) = \int_{[0,1]^2} I_p(f(x,y))dxdy.$$

The following lemma from [11] tells us that it could be extended to $\tilde{\mathcal{W}}$ as well and is a rate function in Definition 1.2.2.

Lemma 2.1.6. *The function I_p is well-defined on $\tilde{\mathcal{W}}$ and is a rate function under the cut metric $\tilde{d_\square}$.*

2.2 Regularity lemma and counting lemma

2.2.1 ϵ-regular partition

For a graph $G = (V, E)$ on n vertices and $X, Y \subset V$ let

$$E_G(X, Y) := \{(u, v) \in E | u \in X, v \in Y\} \text{ and } e_G(X, Y) := |E_G(X, Y)| \tag{2.3}$$

The *edge density* of the pair (X, Y) is defined to be

$$d_G(X, Y) := \frac{e_G(X, Y)}{|X||Y|}. \tag{2.4}$$

For two subsets of vertices A, B, we define a regularity condition under which the density between two large subsets $X \subset A$ and $Y \subset B$ behaves as if the edges are randomly distributed between A and B with density $d_G(A, B)$.

Definition 2.2.1. *For $\epsilon > 0$, a pair of disjoint subsets (A, B) of V are ϵ-regular if for every $X \subset A$ and $Y \subset B$ with $|X| \geq \epsilon|A|$ and $|Y| \geq \epsilon|B|$ we have*

$$|d_G(A, B) - d_G(X, Y)| \leq \epsilon.$$

A generalization of this condition can be put on the partition of V.

Definition 2.2.2. *A partition $\{V_0, ..., V_K\}$ of V is an ϵ-regular partition if*

i. $|V_0| \leq \epsilon n$,

ii. $|V_1| = ... = |V_K|$,

iii. all but at most ϵK^2 of the pairs (V_i, V_j) with $1 \leq i < j \leq K$ are ϵ-regular.

2.2.2 Szemerédi's regularity lemma

The regularity lemma to some extent gives the existence of such regular partitions.

Theorem 2.2.3. *For every $\epsilon > 0$ and integer $m \geq 1$ there exists an integer $M = M(\epsilon, m)$ so that every graph G with $e_G \geq M$ has an ϵ-regular partition $V_0, ..., V_K$ for some $m \leq K \leq M$.*

The regularity lemma has the following implication on the graphon space $(\tilde{\mathcal{W}}, \tilde{d_\square})$.

Theorem 2.2.4. *The metric space $(\tilde{W}, \tilde{d_\square})$ is compact.*

2.2.3 Counting lemma

For a graph H with $V(H) = [k]$ and a collection of subsets $S = \{V_1, ..., V_k\}$ of V, let $\mathrm{Hom}_S(H, G) = \{f \in \mathrm{Hom}(H, G) | f(i) \in V_i \text{ for } V_i \in S\}$. In a lot of situations, the regularity lemma is used together with the following counting lemma after we get an ϵ-regular partition.

Theorem 2.2.5. *For $\epsilon > 0$, if the collection S satisfies (V_i, V_j) is ϵ-regular for all $(i, j) \in E(H)$, then*

$$\left| |Hom_S(H, G)| - \prod_{(i,j)\in E(H)} d(V_i, V_j) \cdot \prod_{i=1}^{k} |V_i| \right| \leq \epsilon \cdot e_H \prod_{i=1}^{k} |V_i|.$$

An implication of the counting lemma is the following continuity of the homomorphism counting function in the graphon space.

Theorem 2.2.6. *A sequence of graphs $\{G_n\}_{n=1}^{\infty}$ converges to $f \in \mathcal{W}$ if and only if $\lim_{n\to\infty} \tilde{d_\square}(\tilde{G}_n, \tilde{f}) = 0$.*

2.3 Main result

Standard large deviation techniques from 1.2 could give a large deviation principle for the weak topology of $\mathcal{W}$. However, the homomorphism counting function t being a non linear function is not continuous with respect to this topology and we could not get an estimate of the upper tail probability. Instead, they established the following large deviation principle on $\tilde{W}$ with respect to the topology given by the cut metric.

Theorem 2.3.1. *For each fixed $p \in (0,1)$, the sequence $\tilde{\mathbb{P}}_{n,p}$ satisfies a large deviation principle in $(\tilde{\mathcal{W}}, \tilde{d}_\square)$. Explicitly, for any closed set $\tilde{F}$ we have*

$$\lim_{n\to\infty} \sup \frac{2}{n^2} \log \tilde{\mathbb{P}}_{n,p}(\tilde{F}) \le - \inf_{\tilde{f}\in\tilde{F}} I_p(\tilde{f}). \tag{2.5}$$

and for every open set $\tilde{U}$ we have

$$\lim_{n\to\infty} \inf \frac{2}{n^2} \log \tilde{\mathbb{P}}_{n,p}(\tilde{U}) \ge - \inf_{\tilde{f}\in\tilde{U}} I_p(\tilde{f}). \tag{2.6}$$

As an example, by the continuity of t in the cut metric given in Theorem 2.2.6 and together with showing the continuity of ϕ in δ, they get a large deviation principle with respect to the triangle count.

Theorem 2.3.2. *For each fixed $p \in (0,1)$ and fixed $\delta > 0$, we have*

$$\zeta(K_3, n, p, \delta) = -(1 + o(1))\phi(K_3, n, p, \delta).$$

2.4 Proof sketch

For $h \in \mathcal{W}$ and $\eta > 0$, let $B(h, \eta) = \{g \in \mathcal{W} : d_\square(h, g) \le \eta\}$ and $\tilde{B}(\tilde{h}, \eta) = \{\tilde{g} \in \tilde{\mathcal{W}} : \tilde{d}(\tilde{h}, \tilde{g}) \le \eta\}$. Recall the map $\tau : h \to \tilde{h}$, let $B(\tilde{h}, \eta) = \tau^{-1}\tilde{B}(\tilde{h}, \eta) \in \mathcal{W}$ be the union of all orbits. By the compactness of $(\tilde{\mathcal{W}}, \tilde{d})$ given in Theorem 2.2.4, it suffices to prove Theorem 2.3.1 for such local balls. For the upper bound, the main strategy is to use regularity lemma to give a covering of $B(\tilde{h}, \eta)$ with finitely many balls that are weakly closed and then apply the standard large deviation principle on these weakly closed balls. For the lower bound, for each ball, they construct a measure concentrated on the ball and then apply the tilting argument. We describe each one in more detail below.

2.4.1 Upper bound

It suffices to show that for every ball $\tilde{B}(\tilde{h}, \eta)$ in $\tilde{\mathcal{W}}$ we have

$$\lim_{\eta\to 0} \lim_{n\to\infty} \sup \frac{2}{n^2} \log \mathbb{P}_{n,p}(B(\tilde{h}, \eta)) \le -I_p(\tilde{h}) \tag{2.7}$$

For a graph $G = (V, E)$ on n vertices and a partition of the vertices $S = \{V_0 \cup V_1 ... V_K\}$ such that $|V_0| = b$ and $|V_1| = ... = |V_K| = a$, we can identify each V_0 with intervals $E_0 = [0, \frac{b}{n}]$ and V_i with $E_i = [\frac{b+(i-1)a}{n}, \frac{b+ia}{n}]$ for $1 \leq i \leq K$. Recall the edge density $d_G(V_i, V_j)$ defined in (2.4), this gives a block graph f_S^G where

$$f_S^G(x, y) = \sum_{i \neq j} d_G(V_i, V_j) \mathbf{1}_{E_i}(x) \mathbf{1}_{E_j}(y).$$

By giving a regular partition, the regularity lemma shows that for every $\epsilon > 0$, there is a bounded $m \leq K \leq M$ such that for every graph G there is a permutation π so that $d_\square(f_S^{\pi G}, f^{\pi G}) \leq \epsilon$. Let $\mathcal{V}_K$ be the set of block functions with K blocks. Thus we can cover the set of graphons with positive measure under $\mathbb{P}_{n,p}$ by $\cup_{K=m}^M B(\mathcal{V}_K, \epsilon)$. So it suffices to show that for each K we have

$$\lim_{\eta \to 0} \lim_{n \to \infty} \sup \frac{2}{n^2} \log \mathbb{P}_{n,p}(B(\tilde{h}, \eta) \cap B(V_K, \epsilon)) \leq -I(\tilde{h}) + o(\epsilon).$$

Since V_K is a compact set of functions we can further cover $B(V_K, \epsilon)$ by a finite number of spheres $B(g, 2\epsilon)$ in $\mathcal{W}$. By a union bound it suffices to bound $\mathbb{P}_{n,p}(B(\tilde{h}, \eta) \cap B(g, 2\epsilon))$. Note that $B(g, 2\epsilon)$ is weakly closed in $\mathcal{W}$, so by Corollary 1.2.10,

$$\lim_{n \to \infty} \sup \frac{2}{n^2} \log \mathbb{P}_{n,p}(B(g, 2\epsilon)) \leq -\inf_{f \in B(g, 2\epsilon)} I_p(f). \tag{2.8}$$

Thus

$$\begin{aligned} \lim_{\eta \to 0} \lim_{n \to \infty} \sup \frac{2}{n^2} \log \mathbb{P}_{n,p}(B(\tilde{h}, \eta) \cap B(g, 2\epsilon)) &\leq \lim_{n \to \infty} \sup \frac{2}{n^2} \log \mathbb{P}_{n,p}(B(g, 2\epsilon)) \\ &\leq -\inf_{f \in B(g, 2\epsilon)} I_p(f) = -I_p(\tilde{h}) + o(\epsilon). \end{aligned} \tag{2.9}$$

The last equality holds since $\tilde{d_\square}(\tilde{g}, \tilde{h}) \leq \eta + 2\epsilon$ assuming $B(\tilde{h}, \eta) \cap B(g, 2\epsilon) \neq \emptyset$ and I_p is lower semi-continuous as shown in Lemma 2.1.6. Then Equation (2.7) would be implied by taking $\epsilon \to 0$.

2.4.2 Lower bound

For given $h \in \mathcal{W}$ and n, consider discretizing h by replacing the value of h in $[\frac{i-1}{n}, \frac{i}{n}] \times [\frac{j-1}{n}, \frac{j}{n}]$ with its average in the block $h_n(i,j)$. Let h_n be the step wise constant function we get. Define the local ball $B_{\epsilon,n} = \{f : d_{\square}(f, h_n) \leq \epsilon\}$, since $d_{\square}(h_n, h) \to 0$ it then suffices to show that

$$\lim_{n\to\infty} \inf \frac{2}{n^2} \log \mathbb{P}_{n,p}(B_{\epsilon,n}) \geq -I_p(h). \tag{2.10}$$

The lower bound is to use a tilting argument by constructing the following measure $\mathbb{P}_{h_n}$ that is concentrated around h_n. Let $\mathbb{P}_{h_n}$ be measure on $\mathcal{W}$ induced by the inhomogeneous random graph where edge (i,j) are connected with probability $h_n(i,j)$. Then a standard concentration inequality gives that $\lim_{n\to} \mathbb{P}_{h_n}(B_{\epsilon,n}) = 1$ and by the tilting argument outlined in section 1.2.1.2 with $\mathbb{P} = \mathbb{P}_{n,p}$ and $\mathbb{Q} = \mathbb{P}_{n,h}$ we have

$$\lim_{n\to\infty} \inf \frac{2}{n^2} \log \mathbb{P}_{n,p}(B_{\epsilon,n}) \geq -\lim_{n\to\infty} \frac{2}{n^2} \int \log \frac{d\mathbb{P}_{n,h}}{d\mathbb{P}_{n,p}} d\mathbb{P}_{n,h} \tag{2.11}$$

The only thing left to do is to simplify the right hand side. Note that

$$\int \log \frac{d\mathbb{P}_{n,h}}{d\mathbb{P}_{n,p}} d\mathbb{P}_{n,h} = \sum_{i>j} \big(h_n(i,j) \log \frac{h_n(i,j)}{p} + (1 - h_n(i,j)) \log \frac{1 - h_n(i,j)}{1-p} \big).$$

By symmetry of h we then have

$$-\lim_{n\to\infty} \frac{2}{n^2} \int \log \frac{d\mathbb{P}_{n,h}}{d\mathbb{P}_{n,p}} d\mathbb{P}_{n,h} = -I_p(h). \tag{2.12}$$

Combine (2.11) and (2.11) we get (2.10).

Chapter 3

Variational reduction in sparse regime

3.1 Introduction

The methods of [11] that we introduced in the previous chapter does not apply to the sparse regime due to the bad quantitative dependencies of the regularity lemma. Specifically, in order to bound the fluctuation of I_p within a ball $B(g, 2\epsilon)$ in Equation (2.9) we need to take very small ϵ. However, the upper bound on the number of balls given by the regularity lemma, which is unfortunately tight, could not give us the desired result. In this chapter we sketch some of the ideas behind [12] and how they resolved this issue. Their method can be seen as a quantitative strengthening of [11]. One of their main points is that instead of covering the space with small weakly closed sets as is done in [11], one can find a covering by convex sets using spectral methods and estimate the probability of each convex set accurately without requiring them to be too small. With this new approach, they are able to prove the following theorem establishing a large deviation principle for the widest range of p currently known for general graphs.

Theorem 3.1.1. *For any graph H, let $\Delta_*(H) = \frac{1}{2}\max_{v_1,v_2\in E(H)}\{\deg_H(v_1) + \deg_H(v_2)\}$. If $p = p(n)$ satisfies $n^{-1}\log n \ll p^{2\Delta_*(H)} \ll 1$ then for any $\delta > 0$,*

$$\zeta(H, n, p, \delta) = -(1 + o(1))\phi(H, n, p, \delta + o(1)).$$

3.2 Matrix preliminaries

Their proof relies heavily on the adjacency matrix and spectral inequalities. For a set Ω, let $\mathrm{Sym}_n(\Omega)$ be the set of symmetric $n \times n$ matrices with entries in Ω and $\mathrm{Sym}_n^0(\Omega)$ those with 0 on the diagonal. Let $\mathrm{Sym}_{n,k}(\mathbb{R})$ be real symmetric matrices of rank at most k. Let $\mathcal{X}_n = \mathrm{Sym}_n^0([0,1])$ and $\mathcal{A}_n = \mathrm{Sym}_n^0(\{0,1\})$. For $X \in \mathrm{Sym}_n(\mathbb{R})$, we order its eigenvalues in non-increasing order of modulus:

$$|\lambda_1(X)| \geq |\lambda_2(X)| \geq ... \geq |\lambda_n(X)|.$$

Recall that the Schatten norm is defined as

$$\|X\|_{S_\alpha} = (\sum_{i=1}^n |\lambda_i(X)|^\alpha)^{1/\alpha}, \alpha \in [1, \infty].$$

It can represent many common matrix norms. We have the spectral norm $|\lambda_1(X)| = \|X\|_{S_\infty}$ and it also equals the $l_2^n \to l_2^n$ operator norm $\|X\|_{\mathrm{op}}$. Moreover, $\|X\|_{S_2}$ equals the Hilbert-Schmidt norm $\|X\|_{\mathrm{HS}} = (\mathrm{Tr}X^2)^{1/2}$ with inner product $\langle X, Y\rangle = \mathrm{Tr}(XY)$. Let

$$\mathbb{B}_{\mathrm{HS}}(r) = \{X \in \mathrm{Sym}_n(\mathbb{R}) : \|X\|_{\mathrm{HS}} \leq r\} \text{ and } \mathbb{B}_{\mathrm{op}}(X, r) = \{Y \in \mathrm{Sym}_n(\mathbb{R}) : \|X - Y\|_{\mathrm{op}} \leq r\},$$

we then have $\mathcal{X}_n \subset \mathbb{B}_{\mathrm{HS}}(n)$.

We can bound the difference in cycle count in terms of matrix norms with the following lemma. Note that for $X \in \mathcal{X}_n$, we have $t(C_{2l}, X) = n^{-2l}\|X\|_{S_{2l}}^{2l}$ and $t(C_l, X) = n^{-l}\mathrm{Tr}X^l$.

Lemma 3.2.1. *For $X_1, X_2 \in \mathcal{X}_n$, we have*

$$|t(C_l, X_1) - t(C_l, X_2)| \leq ln^{-l} \cdot \|X_1 - X_2\|_{op} \cdot (\|X_1\|_{S_{l-1}}^{l-1} + \|X_2\|_{S_{l-1}}^{l-1})$$

Proof. Consider order the eigenvalues directly instead of by modulus, we have that by Weyl's Inequality,

$$|\lambda_i(X_1) - \lambda_i(X_2)| \leq \|X_1 - X_2\|_{\mathrm{op}}.$$

Since $|a^l - b^l| \leq l|a-b|(|a|^{l-1} + |b|^{l-1})$ for any $a, b \in \mathbb{R}$ and $l \in \mathbb{N}$, we have

$$|\mathrm{Tr}X_1^l - \mathrm{Tr}X_2^l| \leq \sum_{i=1}^n |\lambda_i(X_1)^l - \lambda_i(X_2)^l|$$

$$\leq \sum_{i=1}^{n} l|\lambda_i(X_1) - \lambda_i(X_2)| \cdot (|\lambda_i(X_1)|^{l-1} + |\lambda_i(X_2)|^{l-1})$$

$$\leq l\|X_1 - X_2\|_{\text{op}} \cdot (\|X_1\|_{S_{l-1}}^{l-1} + \|X_2\|_{S_{l-1}}^{l-1}).$$

Thus

$$|t(C_l, X_1) - t(C_l, X_2)| = n^{-l}|\text{Tr}X_1 - \text{Tr}X_2| \leq ln^{-l} \cdot \|X_1 - X_2\|_{op} \cdot (\|X_1\|_{S_{l-1}}^{l-1} + \|X_2\|_{S_{l-1}}^{l-1}).$$

□

As a corollary we have

$$|t(K_3, X_1) - t(K_3, X_2)| \leq 3n^{-3}\|X_1 - X_2\|_{\text{op}} \cdot (\|X_1\|_{\text{HS}}^2 + \|X_2\|_{\text{HS}}^2) \leq 6n^{-1} \cdot \|X_1 - X_2\|_{op}. \quad (3.1)$$

3.3 Proof Sketch

Recall I_p defined in (1.1), for a set $\mathcal{E} \subset \mathbb{R}^d$ we define

$$I_p(\mathcal{E}) := \inf\{I_p(x) : x \in \mathcal{E} \cap [0,1]^d\}.$$

For a function $h : [0,1]^d \to \mathbb{R}$, let the super-level set

$$\mathcal{U}(h, \delta) = \{x \in [0,1]^d : h(x) \geq \delta\}.$$

Let μ_p be the product Bernoulli(p) measure on $\mathcal{A}_n$. Fix a graph H, recall that the homomorphism counting function $t_H : \mathcal{A}_n \to \mathbb{R}$, we then have

$$I_p(\mathcal{U}(t_H, (1+\delta)p^{|E(H)|}) = \phi(H, n, p, \delta). \quad (3.2)$$

The goal is to estimate $\mu_p(\mathcal{U}(t_H, (1+\delta)p^{|E(H)|}))$.
For the lower bound, the proof essentially follows similar tilting argument as in section 2.4.2 thus we focus on the upper bound. The strategy is that, instead of covering the space with weakly closed set in the dense case where classical large deviation theory gives an upper bound in Equation (2.8), they cover the space with convex sets where we have the following upper bound.

Proposition 3.3.1 (Equation (4.5.6) and Exercise 2.2.23(b) in [15]). *For any closed convex set* $\mathcal{K} \subset \mathbb{R}^d$,

$$\mu_p(\mathcal{K}) \leq \exp(-I_p(\mathcal{K})).$$

After finding a covering, a union bound would give us the following upper bound.

Proposition 3.3.2. *Let* $h : [0,1]^d \to \mathbb{R}$. *Suppose there is a finite family* $\{\mathcal{B}_i\}_{i\in\mathcal{I}}$ *of closed convex sets in* $\mathbb{R}^d$, *and an "exceptional" set* $\mathcal{E} \in \{0,1\}^d$ *and* $\epsilon > 0$ *such that*

$$\{0,1\}^d \setminus \mathcal{E} \subset \cup_{i\in\mathcal{I}}\mathcal{B}_i \text{ and } \forall i \in \mathcal{I}, \forall x, y \in \mathcal{B}_i, h(x) - h(y) \leq \epsilon.$$

Then for any $p \in [0,1]$ *and* $\lambda \in \mathbb{R}$,

$$\mu_p(\mathcal{U}(h,\lambda)) \leq |\mathcal{I}| \exp\big(-I_p(\mathcal{U}(h,\lambda-\epsilon)\big) + \mu_p(\mathcal{E}).$$

So the more precise proof strategy is to find an efficient covering of $\mathcal{X}_n$ by convex sets $\{\mathcal{B}_i\}$ so that $\mu_p(\mathcal{E})$ is small for the uncovered exceptional set $\mathcal{E}$ and the function t_H does not fluctuate much within each $\mathcal{B}_i$. To achieve this, they establish a spectral regularity lemma and counting lemma inspired from the spectral proof of the classical regularity lemma [19, 31].
A notable feature is that compared to the the cut norm $d_\square$ in the classical regularity lemma on $\mathcal{W}$, they use the operator norm $\|\cdot\|_{op}$ on matrices.

Proposition 3.3.3 (Spectral regularity lemma). *For some absolute constant* $C_* < \infty$, *any* $\delta_0 < 1$, $K, \Delta > 0$, $n^{-1}\log n \leq p \leq 1$ *and* $n \geq k \geq K(p^\Delta/\delta_0^2)\log(1/p)$, *there exists a partition* $\mathcal{A}_n = \cup_{i=1}^N \mathcal{E}_i$ *such that*

$$\log N \leq k(n+2)\log\frac{3n}{\delta_0}, \tag{3.3}$$

$$\mu_p(\mathcal{E}) \leq 4\exp(-Kn^2p^\Delta\log(1/p)), \tag{3.4}$$

and for each $1 \leq i \leq N$, *there exists* $Y_j \in Sym_{n,k}(\mathbb{R}) \cap \mathbb{B}_{HS}(n)$ *with*

$$\max_{X\in\mathcal{E}_i}\{\|X - (p\cdot A(K_n) + Y_j)\|_{op}\} \leq C_*(\sqrt{np} + \delta_0 n), \tag{3.5}$$

recall that K_n *is the clique on* n *vertices.*

Equation (3.3) upper bounds the number of coverings, corresponding to the constant "K" in the original regularity lemma of Theorem 2.2.3. Equation (3.4) upper bounds the probability of the exceptional set $\mathcal{E}$, corresponding to the condition "$|V_0| \leq \epsilon n$" in the regular partition. Equation (3.5) essentially says that each X could be approximated by low rank matrices in spectral norm.

Essentially, their regularity lemma gives a covering $\cup \mathbb{B}_{\mathrm{op}}(Y_i, \epsilon)$ with operator norm of the adjacency matrix. Then a corresponding spectral counting lemma (whose precise statement we defer later to Proposition 3.3.5) like Lemma 3.2.1 gives a bound of $|t_H(X_1) - t_H(X_2)|$ for $X_1, X_2 \in \mathbb{B}_{\mathrm{op}}(Y, \epsilon)$. Combining these two we can apply Proposition 3.3.2 to get the desired upper bound. In the next subsection, we illustrate the main idea with $H = K_3$.

3.3.1 Illustration with K_3.

We illustrate the main idea by proving the following weak form of Theorem 3.1.1 for the case of triangles.

Proposition 3.3.4. *If $p = p(n)$ satisfies $n^{-1} \log n \ll p^8 \ll 1$ then for any $\delta > 0$,*

$$\zeta(K_3, n, p, \delta) = -(1 + o(1))\phi(K_3, n, p, \delta + o(1)).$$

Proof. Let $\epsilon > 0$ be chosen later. Proposition 3.3.3 gives a set $\mathcal{N} \subset \mathcal{X}_n$ of size $O(\epsilon^{-2} n \log n)$ so that for each $X \in \mathcal{X}_n$, there is $Y \in \mathcal{N}$ such that $\|X - Y\|_{op} \leq \epsilon n$. To each $Y \in \mathcal{N}$ we associate $\mathbb{B}_{\mathrm{op}}(Y, \epsilon n)$. For any $X_1, X_2 \in \mathbb{B}_{\mathrm{op}}(Y, \epsilon n)$, by (3.1) and triangle inequality,

$$|t_{K_3}(X_1) - t_{K_3}(X_2)| \leq 6n^{-1} \cdot (\|X_1 - Y\|_{op} + \|X_2 - Y\|_{op}) \leq 12\epsilon.$$

Pick $\epsilon = o(p^3)$, by Proposition 3.3.2 with $\{\mathcal{B}_i\}_{i \in \mathcal{I}} = \{\mathbb{B}_{\mathrm{op}}(Y, \epsilon n)\}_{Y \in \mathcal{N}}$ and $\mathcal{E} = \emptyset$,

$$\mathbb{P}(t(K_3, G) \geq (1 + \delta)p^3) = \mu_p\big(\mathcal{U}(t_{K_3}, (1 + \delta)p^3)\big) \leq |\mathcal{N}| \exp(-I_p(\mathcal{U}(t_{K_3}, (1 + \delta)p^3 - 12\epsilon)))$$

$$= |\mathcal{N}| \exp(-I_p(\mathcal{U}(t_{K_3}, (1 + \delta)p^3 - o(p^3)))) = |\mathcal{N}| \exp(-\phi(K_3, n, p, \delta - o(1))).$$

When $p \gg (\log n/n)^{1/8}$, by Theorem 1.17 the main term dominates, so we have

$$\zeta(K_3, n, p, \delta) = \log \mathbb{P}(t(K_3, G) \geq (1+\delta)p^3) \leq -\phi(K_3, n, p, \delta - o(1)) + \log |\mathcal{N}|$$

$$= -(1 + o(1))\phi(K_3, n, p, \delta - o(1)).$$

$\square$

3.3.2 Spectral counting lemma and refined argument

With a simple version for cycle counts given in Lemma 3.2.1, we now introduce their general spectral counting lemma. For a graph H, we use $F \preccurlyeq H$ to mean that F is an induced subgraph of H(i.e. $F = H[V']$ for some $V' \subset V(H)$) and $F \prec H$ if $F \preccurlyeq H$ and $F \neq H$. Let $\mathcal{L}_H(\lambda) := \{X \in \mathcal{X}_n : t_H(X) \leq \lambda p^{e_H})\}$, the following counting lemma shows that for any H and $K = O(1)$, after localizing to a region $\mathcal{B}$ that is near all sub-level sets $\mathcal{L}_F(K)$ with $F \prec H$, the function t_H is $O_H(p^{e_H - \Delta_*}/n)$-Lipschitz on $(\mathcal{X}_n, \|\cdot\|_{\mathrm{op}})$.

Proposition 3.3.5 (Spectral counting lemma). *Given $p \in (0,1), n \in \mathbb{N}$ and $K \geq 1 \geq \epsilon_0$, for any graph H with $\Delta_*(H) \leq \Delta_*$ and any convex set $\mathcal{B} \subset \mathcal{X}_n$ satisfying*

$$\mathcal{B} \cap \mathcal{L}_F(K) \neq \emptyset, \forall F \prec H,$$

$$\sup_{X,Y \in \mathcal{B}} \{\|X - Y\|_{op}\} \leq \epsilon_0 n p^{\Delta_*},$$

we have that for all $F \preccurlyeq H$,

$$\sup_{X,Y \in \mathcal{B}} \{|t_F(X) - t_F(Y)|\} \leq C_H \epsilon_0 K$$

for a constant $C_H < \infty$ depending only on H.

Note that in our illustration with K_3, we take the exceptional set $\mathcal{E}$ to be empty. In their refined argument, they take $\mathcal{E}$ to consist of matrices with too many outlier eigenvalues to get a wider range of p.

Chapter 4

Combinatorial reduction via entropic stability

4.1 Introduction

A framework for proving sharp asymptotics for the upper tail of general polynomials of Bernoulli random variables was introduced in [21]. Let $x = (x_1, ..., x_n)$ where the x_is are independent $\mathrm{Ber}(p)$. For a random variable $X = f(x)$, they defined a combinatorial optimization problem

$$\Phi_X(\delta) = \min\{|I| \log(1/p) : I \subset [n] \text{ and } \mathbb{E}_I[X] \geq (1+\delta)\mathbb{E}[X]\}. \tag{4.1}$$

where $\mathbb{E}_I[X] = \mathbb{E}[X|x_I = 1]$. Recalling the upper tail for subgraph counts R in (1.8), they proved the following theorem on reducing the upper tail to the combinatorial optimization problem Φ in (4.1).

Theorem 4.1.1 (Theorem 1.5 in [21])**.** *For every $\Delta \geq 2$, every connected, nonbipartite, Δ-regular graph H and all positive real number δ, suppose that we have $n^{-1}(\log n)^{v_H^2\Delta} \ll p^{\Delta/2} \ll 1$, then we have*

$$R(H, \delta, n, p) = -(1 + o(1))\Phi(H, \delta + o(1), n, p).$$

Additionally, if $p^{\Delta/2} \gg n^{-1/2-o(1)}$, then the nonbipartite assumption is not necessary.

Remark 4.1.2. *They showed that Φ and the variational optimization problem ϕ has the* *same asymptotic solution. It is clear that $\Phi \geq \phi$ since the core graph G can be extended to an element $\mathcal{G}$ in $\mathcal{G}_n$ by setting the weights of edges outside G to be p. We then have $e_G \log(1/p) = I_p(\mathcal{G})$ and $|Hom(H, \mathcal{G})| \asymp \mathbb{E}_G[N(H, G')]$. The lower bound is given by the clique and hub construction in Chapter 5.*

Just like the works of [11] and [12] introduced in Chapters 2 and 3, their method is also based on covering. The difference is that instead of covering the whole space $\mathcal{X}_n$ they get a more efficient covering of the large deviation events with certain "near optimizers" of Φ, which they call *cores*. Just like the union bound used in [11] and [12] they showed that the random variable X satisfies a large deviation principle with rate function Φ if the number of cores have a certain upper bound, a property they call *entropic stability.* They proved the following theorem which we will apply later in Chapter 6 when proving our result about hypergraphs.

Theorem 4.1.3 (Theorem 3.1 in [21])**.** *For every integer d and all positive real numbers ϵ and δ with $\epsilon < 1/2$. there is a positive $K = K(d, \epsilon, \delta)$ such that the following holds. Let Y be a sequence of N independent Ber(p) random variables for some $p \in (0, 1-\epsilon]$ and let $X = X(Y)$ be a nonzero polynomial with nonnegative coefficients and total degree at most d such that $\Phi_X(\delta - \epsilon) \geq K \log(1/p)$. Denote by $\mathcal{I}^*$ the collections of all subsets of $I \subset [N]$ such that*

(C1) $\mathbb{E}_I[X] \geq (1 + \delta - \epsilon)\mathbb{E}[X]$,

(C2) $|I| \leq K \cdot \Phi_X(\delta + \epsilon)$, *and*

(C3) $\min_{i \in I}(\mathbb{E}_I[X] - \mathbb{E}_{I \setminus \{i\}}[X]) \geq \mathbb{E}[X]/(K \cdot \Phi_X(\delta + \epsilon))$

and assume that for every $m \in \mathbb{N}$, there are at most $(1/p)^{\epsilon m/2}$ sets of size m in $\mathcal{I}^*$. Then

$$(1 - \epsilon)\Phi_X(\delta - \epsilon) \leq -\log \mathbb{P}(X \geq (1 + \delta)) \leq (1 + \epsilon)\Phi_X(\delta + \epsilon)$$

Remark 4.1.4. *The elements of $\mathcal{I}^*$ are referred to as cores and the property $|\{I \subset \mathcal{I}^* : |I| = m\}| \leq (1/p)^{\epsilon m/2}$ is called entropic stability.*

Based the work of [21], [2] further improved this to all cycles C_l. To verify that a random variable satisfies the entropic stability property is problem specific and in order to do so for the subgraph count problem, they use many non trivial results from graph theory. In the next section, we are going to sketch how their framework is applied to the subgraph count problem.

4.2 Notations

Let $\mathrm{Emb}(H,G) = \{f \in \mathrm{Hom}(H,G), f \text{ is injective}\}$ be the set of graph embeddings and $\mathrm{Emb}(H,G;e,e')$ be the set of embeddings that maps the edge e of H to the edge e' of G. For vertices $u,v \in V(G)$, $deg_G(u,v)$ is the number of vertices w so that $(u,w),(v,w) \in E(G)$.

4.3 Illustration with K_3

In this section, we give a self-contained illustration of the main idea with $H = K_3$. Essentially, Theorem 4.1.3 established a general framework relating upper tail probability to the existence of cores (see Claim 1 below and the proof for general Bernoulli random variables is similar). To apply such framework, the main task is to bound the number of cores (see Claim 2 below), which is problem specific. Letting X be the random variable $N(K_3, G_{n,p})$, we prove the following.

Proposition 4.3.1. *Assume that $n^{-1}\log n \ll p \ll 1$, for every $\delta > 2\epsilon > 0$ and sufficiently large n,*

$$(1-\epsilon)\Phi_X(\delta-\epsilon) \le -\log \mathbb{P}(X \ge (1+\delta)\mathbb{E}[X]) \le (1+\epsilon)\Phi_X(\delta+\epsilon).$$

The upper bound is by the construction of cliques and hubs in Chapter 5 and here we focus on the lower bound. Let $C = C(\epsilon,\delta)$ be a constant chosen later. Similar to Theorem 4.1.3, a graph $G^* \subset K_n$ is a *core* if

(C1) $\mathbb{E}_{G^*}[X] \ge (1+\delta-2\epsilon)\mathbb{E}[X]$,

(C2) $e_{G^*} \le Cn^2p^2\log(1/p)$,

(C3) $\min_{e\in E(G^*)}(\mathbb{E}_{G^*}[X] - \mathbb{E}_{G^*\setminus e}[X]) \ge \epsilon\mathbb{E}[X]/(Cn^2p^2\log(1/p))$.

The proof strategy is to first upper bound the upper tail probability by the existence of a core and then bound the probability of a core using entropic stability. Specifically we prove the following two claims.

Claim 1. $\mathbb{P}(X \geq (1+\delta)\mathbb{E}[X]) \leq (1+\epsilon)\mathbb{P}(G_{n,p}$ contains a core).

Claim 2. For every m, there are at most $(1/p)^{\epsilon m}$ cores with m edges.

First we see how these two claims will yield the upper bound. We have

$$\mathbb{P}(X \geq (1+\delta)\mathbb{E}[X]) \leq (1+\epsilon)\mathbb{P}(G_{n,p} \text{ contains a core})$$

$$\leq (1+\epsilon)\sum_{m=0}^{\infty} p^m \cdot |\{G^* \subset K_n : e_{G^*} = m\}|$$

$$\leq (1+\epsilon)\sum_{m=m_{min}}^{\infty} p^{(1-\epsilon)m},$$

where m_{min} is the minimal number of edges a core can have. (C1) implies $\Phi_X(\delta - 2\epsilon) \leq m_{min} \cdot \log(1/p)$, together with the assumption $p \ll 1$ gives

$$\mathbb{P}(X \geq (1+\epsilon)\mathbb{E}[X]) \leq (1+\epsilon)\exp(-(1-2\epsilon)\Phi_X(\delta - 2\epsilon)).$$

Finally, since $\Phi_X(\delta - 2\epsilon) \gg 1$,

$$-\log \mathbb{P}(X \geq (1+\delta)\mathbb{E}[X]) \geq (1-2\epsilon)\Phi_X(\delta - 2\epsilon),$$

thus proving the lower bound with 2ϵ instead of ϵ. Now we prove the two Claims.

Proof of Claim 1

The idea is to start from the following simpler *seed* graph which is easier to relate to the upper tail probability and arrive at a core graph by gradually removing edges from the seed graph. A graph $G \subset K_n$ is a seed if

(S1) $\mathbb{E}_G[X] \geq (1+\delta-\epsilon)\mathbb{E}[X]$

(S2) $e_G \leq Cn^2p^2\log(1/p)$.

First we show that every seed contains a core. Let G be any seed and let $G_0 = G$. For each G_i, let $G_{i+1} = G_i \setminus e$ for some $e \in G_i$ such that $\mathbb{E}_{G_i}[X] - \mathbb{E}_{G_i \setminus e}[X] < \epsilon\mathbb{E}[X]/(Cn^2p^2\log(1/p))$ as long as such e exists. Suppose this process stop at $G^* = G_s$. We now have that G^* is a core. (C3) is clearly satisfied by construction and (C2) is satisfied since $e_{G^*} \leq e_G \leq Cn^2p^2\log(1/p)$. For (C1), since $s \leq e_G \leq Cn^2p^2\log(1/p)$, we have

$$\mathbb{E}_G[X] - \mathbb{E}_{G^*}[X] = \sum_{i=0}^{s-1}(\mathbb{E}_{G_i}[X] - \mathbb{E}_{G_{i+1}}[X]) < \epsilon\mathbb{E}[X].$$

Thus $\mathbb{E}_G[X] \geq (1+\delta-2\epsilon)\mathbb{E}[X]$ and (C1) is satisfied as well. To prove Claim 1, now it suffices to show that

$$\mathbb{P}(X \geq (1+\delta)\mathbb{E}[X]) \leq (1+\epsilon)\mathbb{P}(G_{n,p} \text{ contains a seed}). \tag{4.2}$$

To show this, they employed a high moment argument. This is a refinement of the classical arguments in [22]. Let $Z = \mathbf{1}_{G_{n,p} \text{ does not contain a seed}}$ and let $l = \lceil (C/3)n^2p^2\log(1/p)\rceil$. By Markov's Inequality

$$\mathbb{P}(X \geq (1+\delta), G_{n,p} \text{ contains no seed}) = \mathbb{P}(XZ \geq (1+\delta)\mathbb{E}[X]) \leq \frac{\mathbb{E}[X^lZ]}{(1+\delta)^l\mathbb{E}[X]^l}. \tag{4.3}$$

To upper bound the right hand side, we show by induction that $\mathbb{E}[X^lZ] < (1+\delta-\epsilon)^l\mathbb{E}[X]^l$. Writing X as $X = \sum_T Y_T$ where the sum ranges over all triangles T in K_n and $Y_T = 1_{\{T\subset G_{n,p}\}}$. For every $G \subset K_n$, let $Z_G = 1_{\{G\cap G_{n,p} \text{ does not contain a seed}\}}$. We have $Z_{G'} \leq Z_G$ if $G \subset G'$. Since $Z = Z_{K_n}$, we have for every $k \in [l]$,

$$\mathbb{E}[X^kZ] = \sum_{T_1,\ldots,T_k}\mathbb{E}[Y_{T_1}...Y_{T_k}\cdot Z] \leq \sum_{T_1,\ldots,T_k}\mathbb{E}[Y_{T_1}...Y_{T_k}\cdot Z_{T_1\cup\ldots\cup T_k}]$$

$$\leq \sum_{T_1,\ldots,T_{k-1}}\mathbb{E}[Y_{T_1}...Y_{T_{k-1}}\cdot Z_{T_1\cup\ldots\cup T_{k-1}}]\cdot\sum_{T_k}\mathbb{E}[Y_{T_k}|Y_{T_1}...Y_{T_{k-1}}\cdot Z_{T_1\cup\ldots\cup Tk-1}=1].$$

The key observation here is that we only need to sum over $T_1,...,T_{k-1}$ such that the event $\{Y_{T_1}...Y_{T_{k-1}}\cdot Z_{T_1\cup\ldots\cup T_{k-1}} = 1\}$ has positive probability, equivalent to saying that $T_1\cup...\cup T_{k-1}$ does not contain a seed. Let $H_k = T_1\cup...\cup T_k$, we then have

$$\sum_{T_k}\mathbb{E}[Y_{T_k}|Y_{T_1}...Y_{T_{k-1}}\cdot Z_{T_1\cup\ldots\cup T_{k-1}}=1] = \sum_{T_k}\mathbb{E}[Y_{T_k}|Y_{T_1}...Y_{T_{k-1}}=1] = \mathbb{E}_{H_{k-1}}[X].$$

Note that H_{k-1} does not contain a seed and

$$e_{H_{k-1}} \leq 3(k-1) \leq 3(l-1) \leq Cn^2p^2\log(1/p)$$

thus $\mathbb{E}_{H_{k-1}}[X] < (1+\delta-\epsilon)\mathbb{E}[X]$ and we have the following recursive expression

$$\mathbb{E}[X^k Z] \leq \sum_{T_1,\ldots,T_k} \mathbb{E}[Y_{T_1}\ldots Y_{T_k}\cdot Z_{T_1\cup\ldots\cup T_k}] \leq (1+\delta-\epsilon)\mathbb{E}[X]\cdot \sum_{T_1,\ldots,T_{k-1}} \mathbb{E}[Y_{T_1}\ldots Y_{T_{k-1}}\cdot Z_{T_1\cup\ldots\cup T_{k-1}}],$$

which gives $\mathbb{E}[X^l Z] < (1+\delta-\epsilon)^l \mathbb{E}[X]^l$. Substituting this back to (4.3) we have

$$\mathbb{P}(X \geq (1+\delta) \text{ and } G_{n,p} \text{ contains no seed}) \leq (\frac{1+\delta-\epsilon}{1+\delta})^l.$$

On the other hand,

$$\epsilon\mathbb{P}(G_{n,p} \text{ contains a seed}) \geq \epsilon\exp(-\Phi_X(\delta-\epsilon)) \geq \exp(-(1+\delta)^{2/3}n^2p^2\log(1/p)/2)$$

$$\geq (\frac{1+\delta-\epsilon}{1+\delta})^{(C/3)n^2p^2\log(1/p)} \geq (\frac{1+\delta-\epsilon}{1+\delta})^l.$$

Combining the two bounds yields Claim 1.

Proof of Claim 2

Note that condition (C3) tells us that each edge of a core graph appears in a significant amount of triangle embeddings. This condition will imply a high degree condition on the vertices of G^*. By the bound on the number of high degree vertices, we can have a bound on the number of possible core graphs. Let G^* be a core graph and e be an arbitrary edge of G^*. We then have

$$\epsilon(1-o(1))n^3p^3/(6Cn^2p^2\log(1/p)) \leq \epsilon\mathbb{E}[X]/(Cn^2p^2\log(1/p)) \leq$$

$$\mathbb{E}_{G^*}[X] - \mathbb{E}_{G^*\setminus e}[X] \leq (N(K_3,G^*;e) + N(K_{1,2},G^*;e)\cdot p + np^2)\cdot(1-p),$$

which implies that

$$\frac{\epsilon np}{7C\log(1/p)} \leq N(K_3,G^*;e) + N(K_{1,2},G^*;e)\cdot p + np^2.$$

Thus either

$$N(K_3,G^*;e) \geq \frac{\epsilon np}{15C\log(1/p)} \text{ or } N(K_{1,2},G^*;e) \geq \frac{\epsilon n}{15C\log(1/p)}. \tag{4.4}$$

Let A be the set of vertices of G^* with degree at least $\epsilon np/(30C\log(1/p)))$ and B be the set of vertices with degree at least $\epsilon n/(30C\log(1/p))$. Equation (4.4) implies that every edge of G^* is either contained in A or has an endpoint in B. Note that

$$|A| \le a := \frac{60Cm\log(1/p)}{\epsilon np} \text{ and } |B| \le b := \frac{60Cm\log(1/p)}{\epsilon n}$$

Thus the number of cores with m edges is at most

$$\binom{n}{a}\binom{n}{b}\binom{a^2/2+bn}{m} \le (1/p)^{\epsilon m}$$

for large enough C.

4.4 Entropic stability of $N(H, G_{n,p})$

In this section we illustrate how the entropic stability property could be established for more general graphs using the example $H = K_k$. Note that when letting f to be the graph counting function for H in $G_{n,p}$, (4.1) is exactly $\Phi(H, \delta, n, p)$ in (1.15) and a core I corresponds to a core subgraph G^*. In order to apply Theorem 4.1.3, it suffices to verify the entropic stability property, that is to bound the number of core graphs G^*. In order to achieve this, they showed some limitations on the structure of G^*. They proved that every edge $(u, v) \in E(G^*)$ has to satisfy certain large degree condition. Specifically, they showed the following claim.

Claim 4.4.1. *There exists a positive constant $\eta = \eta(\delta, H, K)$ such that for every core subgraph G^* and every edge $(u, v) \in E(G^*)$, we have that either*

$$deg_{G^*}(u,v) \ge \frac{\eta np^{(k-1)/2}}{(\log(1/p))^{1/(k-2)}} \text{ or } deg_{G^*}(u) + deg_{G^*}(v) \ge \frac{\eta n}{\log(1/p)}.$$

From this claim we are able to establish that G^* has to come from a small set of large degree vertices, thus bounding the number of cores effectively. Now we sketch how they proved this claim.

Note that condition $(C3)$ in Theorem 4.1.3 tells us that each vertex has lots of embeddings containing it. By a simple averaging argument and the upper bound of Φ using the construction in Section 5.2, we have that for each edge (u, v), there exists a nonempty subgraph $J \subset K_k$ with no isolated vertices, an edge $ab \in E(J)$ and a constant $\gamma = \gamma(r, \delta)$

such that

$$|\mathrm{Emb}(J, G^*; ab, uv)| \geq \gamma \cdot n^{v_J - 2} p^{e_J - r + 1} / \log(1/p). \tag{4.5}$$

The next step is to upper bound the left hand side using $deg_{G*}(u, v)$. The following classical result is due to Alon [1].

Lemma 4.4.2. *Let C_l denote the cycle of length l. For every $l \geq 3$ and every graph G,*

$$|Emb(C_l, G)| \leq (2e_G)^{l/2}.$$

Let J_{ab} be the induced graph after removing edges incident to a, b, we have the following graph decomposition results.

Definition 4.4.3. *A fractional independent set of a graph J is an assignment $\alpha : V(J) \to [0, 1]$ that satisfy $\alpha_u + \alpha_v \leq 1$ for every edge $(u, v) \in E(J)$. The fractional independence number, denoted by α^*_J is the maximum of $\sum_{v \in V(J)} \alpha_u$ among all fractional independent sets of J.*

Lemma 4.4.4. *For every graph J with fractional independence number α^*_J, it admits a partition $V(J) = V_1 \cup V_2$ such that $|V_1|/2 + |V_2| = \alpha^*_J$ and V_1 can be covered by vertex disjoint edges and cycles of J.*

In order to provide an upper bound, they first decomposed J_{ab} into disjoint edges and cycles with the Lemma 4.4.4 and then used Lemma 4.4.2. Specifically, choose a partition of J_{ab} that maximizes the cardinality of the set $X = \{c \in V_2 : ac, bc \in E(J)\}$. Let $v_1 = |V_1|, v_2 = |V_2|$ and $x = |X|$, we then have

$$|\mathrm{Emb}(J, G^*; ab, uv)| \leq 2 \cdot deg_{G^*}(u, v)^x \cdot (e_{G^*})^{v_1/2} \cdot \min\{e_{G^*}, n\}^{v_2 - x}. \tag{4.6}$$

Combine (4.5) and (4.6) we will be able to get a lower bound on the degrees of u, v after some other graph analysis that we skip here.

Chapter 5

Solving the variational problem

5.1 Introduction

In this chapter we outline the work of [3] on solving the variational problem in the sparse regime. In Equation (1.17) of Chapter 1 we introduced the solution given by [29] when $H = K_3$. [3] extended this and solved the variational problem for a general graph H with $\Delta(H) \geq 2$ (note that the case of $\Delta(H) = 1$ is nothing but the large deviation for the sum of Bernoulli random variables treated in Example 1.2.1). The result for general H surprisingly involves the following notion of independence polynomial.

Definition 5.1.1. *The independence polynomial of a graph H is defined to be*

$$P_H(x) := \sum_m i_H(m)x^m$$

where $i_H(m)$ is the number of m-element independent sets in H.

Definition 5.1.2. *For a graph H with maximum degree Δ, let $S = \{v \in V(H) : \deg(v) = \Delta\}$, define $H^* := H[S]$ to be the induced graph on these max degree vertices.*

Letting $\theta(H, \delta)$ be the unique positive solution to $P_{H^*}(\theta) = 1+\delta$, this gives the leading constant in the upper tail for $t(H, G_{n,p})$.

Theorem 5.1.3. *Let H be a fixed connected graph with maximum degree $\Delta \geq 2$. For* *any fixed* $\delta > 0$ *and* $n^{-1/\Delta} \ll p \ll 1$,

$$\lim_{n\to\infty} \frac{\phi(H,n,p,\delta)}{n^2 p^\Delta \log(1/p)} = \begin{cases} \min\{\theta(H,\delta), \frac{1}{2}\delta^{2/v_H}\} & \text{if } H \text{ is regular,} \\ \theta(H,\delta) & \text{if } H \text{ is irregular.} \end{cases} \tag{5.1}$$

We partly explain the appearance of the independence polynomial in the next section by showing the upper bound construction.

5.2 Upper bound construction

The upper bound can be proved by exhibiting an explicit weighted graph $G \in \mathcal{G}_n$ with $t(H,G) \geq (1+\delta)p^{e_H}$ and an upper bound on $I_p(G)$ matching the expressions in Equation (5.1). There are two constructions, *clique* and *hub*.

5.2.1 Clique construction for regular graph

Suppose H is Δ-regular and $n^{-2/\Delta} \ll p \ll 1$. Starting with all entries set to p , modify G by setting $a_{ij} = 1$ whenever $i, j \leq s$ for some integer $s \sim \delta^{1/v_H} p^{\Delta/2} n$. Then

$$I_p(G) \sim \frac{1}{2} s^2 I_p(1) \sim \frac{1}{2}\delta^{2/v_H} p^\Delta \log(1/p).$$

Now we show that $t(H,G) \sim (1+\delta)p^{e_H}$. Sum over all $S \subset V(H)$ that gets mapped to $\{1, ..., s\}$ and hide the $1 + o(1)$ factor from those $f : V(H) \to V(G)$ that send two adjacent vertices of H to the same vertex in G, we have

$$t(H,G) \sim n^{-v_H} \sum_{S\subset V(H)} s^{|S|}(n-s)^{v_H-|S|} p^{e_H - e_{H[S]}}$$

$$= \sum_{S\subset V(H)} (\frac{s}{n})^{|S|}(1-\frac{s}{n})^{v_H-|S|} p^{e_H-e_{H[S]}} \sim \sum_{S\subset V(H)} (\delta^{1/v_H} p^{\Delta/2})^{|S|} p^{e_H - e_{H[S]}}$$

Since H is regular and connected, $|S|\Delta/2 - e_{H[S]} > 0$ for all $S \neq \emptyset, V(H)$, in which cases they contribute to a $o(p^{e_H})$ term. Thus $t(H,G) \sim (1+\delta)p^{e_H}$.

By a slight modification of $s \sim \delta^{1/v_H} p^{\Delta/2} n$ we get $t(H,G) \geq (1+\delta)p^{e_H}$, thereby showing $\phi(H,n,p,\delta) \leq (1+o(1))\frac{1}{2}\delta^{2/v_H} p^\Delta \log(1/p)$.

5.2.2 Hub construction for general graph

Suppose H has maximum degree Δ and $n^{-1/\Delta} \ll p \ll 1$. For $\delta > 0$, let $\theta = \theta(H, \delta)$. Starting with $G \equiv p$,modify G by setting $a_{ij} = 1$ whenever $i \leq s$ or $j \leq s$ for some integer $s \sim \theta p^{\Delta} n$. Then

$$I_p(G) \sim sn I_p(1) \sim \theta n^2 p^{\Delta} \log(1/p).$$

Now we show that $t(H, G) \sim (1+\delta)p^{e_H}$. Sum over all subset $S \subset V(H)$ that gets mapped to $\{1, ..., s\}$, we have

$$t(H, G) \sim \sum_{S \subset V(H)} (\frac{s}{n})^{|S|}(1 - \frac{s}{n})^{v_H - |S|} p^{e_{H[V \setminus S]}} \sim \sum_{S \subset V(H)} \theta^{|S|} p^{\Delta |S| + e_{H[V \setminus S]}}$$

Recalling the induced graph H^* defined in Definition 5.1.2 and $e_H(X, Y)$ defined in (2.3), we have

$$\Delta |S| + e_{H[V \setminus S]} \geq e_{H[S]} + e_H(S, V \setminus S) + e_{H[V \setminus S]} = e_H$$

where equality holds only when S is an independent subset of H^*. Thus

$$t(H, G) \sim \sum_{S \text{ is an independent set of } H^*} \theta^{|S|} p^{e_H} = P_{H^*}(\theta) p^{e_H} = (1+\delta) p^{e_H}.$$

By a slight modification of $s \sim \theta p^{\Delta} n$ we have $t(H, G) \geq (1+\delta)p^{e_H}$, thereby showing $\phi(H, n, p, \delta) \leq (1 + o(1))\theta n^2 p^{\Delta} \log(1/p)$.

Letting $a \sim \delta^{1/v_H}$ and $b \sim \theta p^{\Delta}$, the following figure is the graphon interpretation of the clique and hub constructed above, with the blue region being 1 and the orange region being p.

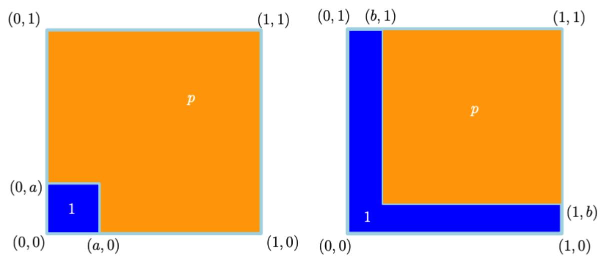

Figure 5.1 The clique graphon and hub graphon (Figure 3 in [3])

5.3 Lower bound proof

In this section, we sketch the proof of the lower bound of ϕ using the example of $H = K_3$. The main idea is to first reformulate ϕ as a continuous optimization problem in the graphon space and then apply the generalized Hölder's inequality from functional analysis.

5.3.1 Graphon reformulation

Recall the space of graphon $\tilde{\mathcal{W}}$ in Definition 2.1.5, the variational problem ϕ can be extended as follows.

Definition 5.3.1. *For a graph $H, \delta > 0$ and $0 < p < 1$, let*

$$\phi(H,p,\delta) = \inf\{\frac{1}{2}\mathbb{E}[I_p(W)] : W \in \tilde{\mathcal{W}}, t(H,W) \geq (1+\delta)p^{e_H}\},$$

where $\mathbb{E}[f(W)] := \int_{[0,1]^2}\int f(W(x,y))dxdy$.

$\phi(H,n,p,\delta)$ and $\phi(H,p.\delta)$ can be related by the following.

Lemma 5.3.2. *For any n, we have $\phi(H,n,p,\delta) \geq n^2\phi(H,p,\delta)$.*

Thus it suffices to establish a lower bound on $\phi(H,p,\delta)$. Specifically, we show that

$$\phi(K_3,p,\delta) \geq (1-o(1))\min\{\frac{1}{2}\delta^{2/3}, \frac{1}{3}\delta\}p^2\log(1/p). \tag{5.2}$$

To prove this inequality, a key ingredient is the following Finner's inequality.

5.3.2 Finner's inequality

The following generalization of Hölder's inequality is due to Finner in [17].

Theorem 5.3.3 (Finner's generalized Hölder's inequality)**.** *Let $\mu_1, ..., \mu_n$ be probability measures on $\Omega_1, ..., \Omega_n$ respectively and let $\mu = \prod_{i=1}^n \mu_i$. For $A \subset [n]$, let $\mu_A = \prod_{j\in A}\mu_j$ and $\Omega_A = \prod_{j\in A}\Omega_j$. Let $A_1, ..., A_m$ be non-empty subsets of $[n]$. Let $f_i \in L^{p_i}(\Omega_{A_i}, \mu_{A_i})$ for each $i \in [m]$, and further suppose that $\sum_{i:j\in A_i} 1/p_i \leq 1$ for all $j \in [n]$. Then*

$$\int \prod_{i=1}^m |f_i| d\mu \leq \prod_{i=1}^m (\int |f_i|^{p_i} d\mu_{A_i})^{1/p_i}. \tag{5.3}$$

The main proof idea is to first relate the left hand side of (5.3) to the homomorphism counting function. To see the connection, let H be a graph on n vertices , each Ω_i be the unit interval and μ_i be the Lebesgue measure and let $\{A_1, ..., A_m\}$ correspond to $E(H)$. Let $f_1 = ... = f_n$ be the graphon W. Then the left hand side of (5.3) becomes $t(H, W)$. Let Δ be the maximum degree of H and let $p_i = \Delta$ for each i. For each j, we have

$$\sum_{i:j\in A_i} \frac{1}{p_i} = \frac{deg(j)}{\Delta} \leq 1.$$

Thus (5.3) becomes

$$t(H, W) \leq \mathbb{E}[W^{\Delta}]^{e_H/\Delta}. \tag{5.4}$$

Note that this can be seen as a continuous analog of Lemma 4.4.2 for cycle counts and Theorem 6.3.2 for general hypergraph counts. The second step is to lower bound $\mathbb{E}[I_p]$ with the right hand side of (5.4). In the case of $H = K_3$, $\Delta = 2$ and we require a quadratic approximation of I_p.

5.3.3 Quadratic approximation of I_p

We now introduce some properties of I_p.

Property 1. I_p is convex and decreasing from 0 to p and increasing from p to 1.

Thus we can assume $W \geq p$. Let $U := W - p$. Then

$$0 \leq U \leq 1 - p \text{ and } t(K_3, p + U) \geq (1 + \delta)p^3.$$

Property 2. There is some $p_0 > 0$ such that for every $0 < p \leq p_0$,

$$I_p(p + x) \geq x^2 I_p(1 - 1/\log(1/p)) \sim x^2 I_p(1) \text{ for any } 0 \leq x \leq 1 - p.$$

This essentially gives a quadratic approximation of $I_p(p + x)$ and implies that

$$\mathbb{E}[I_p(W)] = \mathbb{E}[I_p(p + U)] \geq (1 - o(1)) \int_{[0,1]^2} U(x, y)^2 dxdy \log(1/p). \tag{5.5}$$

5.3.4 Degree thresholding

With the above quadratic approximation, now we shift our focus to lower bound $\mathbb{E}[U^2]$. The idea is to perform a degree threshold and bound the terms above and below the threshold individually. The benefit of this is to account for the contributions coming from the clique and the hub so that the Hölder's inequality turns out to be tight for each summand. For $x \in [0,1]$, let its *normalized degree* be $d(x) := \int_0^1 U(x,y)dy$ and for $b \in (0,1]$, let $B_b := \{x : d(x) \geq b\}$ be the set of points with high degree. By the upper bound construction in the previous section, it suffices to consider U such that $\mathbb{E}[I_p(p+U)] \lesssim p^2 I_p(1)$. We have the following property for such U.

Lemma 5.3.4. *Let U be a graphon satisfying $\mathbb{E}[I_p(p+U)] \lesssim p^2 I_p(1)$. Then*

$$\mathbb{E}[U] \lesssim p^{3/2}\sqrt{\log(1/p)}, \tag{5.6}$$

and

$$\mathbb{E}[U^2] \lesssim p^2, \tag{5.7}$$

furthermore, if $p = o(b)$ then

$$\lambda(B_b) \lesssim \frac{p^2}{b}, \tag{5.8}$$

where λ denotes the Lebesgue measure. Finally,

$$\int_{\bar{B}_b} d(x)^2 dx \lesssim p^2 b. \tag{5.9}$$

5.3.5 Proof of (5.2)

With the above preliminaries, we are ready to prove (5.2) and conclude the formula for $\phi(K_3, n, p, \delta)$ in the sparse regime.

Proof of (5.2)*.* Expanding $t(K_3, W)$ we have,

$$t(K_3, W) - p^3 = t(K_3, U) + 3pt(K_{1,2}, U) + 3p^2\mathbb{E}[U] \geq \delta p^3.$$

By (5.6) $\mathbb{E}[U] = o(p)$, thus

$$t(K_3, U) + 3pt(K_{1,2}, U) \geq (\delta - o(1))p^3. \tag{5.10}$$

By (5.8), for any $p \ll b \ll 1-p$ we have

$$\int_{B_b \times B_b} U(x,y)^2 dxdy \leq \lambda(B_b)^2 \lesssim p^4/b^2 \ll p^2. \tag{5.11}$$

By (5.6) and (5.8), for $\sqrt{p\log(1/p)} \ll b \ll 1$,

$$\int_{[0,1]^3} U(x,y)U(y,z)U(x,z)1\{x \in B_b \text{ or } y \in B_b \text{ or } z \in B_b\}dxdydz \leq 3\lambda(B_b)\mathbb{E}[U] \ll p^3. \tag{5.12}$$

Let

$$\theta_b := p^{-2}\int_{B_b \times \bar{B}_b} U(x,y)^2 dxdy \text{ and } \eta_b := p^{-2}\int_{\bar{B}_b \times \bar{B}_b} U(x,y)^2 dxdy.$$

By (5.12) and Theorem 5.3.3 we have

$$t(K_3,U) = \int_{\bar{B}_b \times \bar{B}_b \times \bar{B}_b} U(x,y)U(y,z)U(x,z)dxdydz + o(p^3)$$

$$\leq (\int_{\bar{B}_b \times \bar{B}_b} U(x,y)^2 dxdy)^{3/2} + o(p^3) = (\eta_b^{3/2} + o(1))p^3.$$

Similarly, by (5.9) and (5.11) and Cauchy-Schwarz inequality, for any $p \ll b \ll 1$,

$$t(K_{1,2},U) = \int_{B_b \times \bar{B}_b \times \bar{B}_b} U(x,y)U(x,z)dxdydz + o(p^2) \leq (\theta_b + o(1))p^2.$$

Combining the above two inequalities with (5.10) we have

$$3\theta_b + \eta_b^{3/2} \geq \delta - o(1).$$

Thus by (5.5),

$$\mathbb{E}[I_p(W)] = \mathbb{E}[I_p(U)] \geq (1-o(1))(\theta_b + \frac{1}{2}\eta_b)p^2\log(1/p)$$

$$\geq (1-o(1)) \min_{x,y\geq 0,\ 3x+y^{3/2}\geq\delta} (x + \frac{1}{2}y)p^2\log(1/p) \sim \min\{\frac{1}{2}\delta^{3/2}, \frac{1}{3}\delta\}p^2\log(1/p).$$

□

To prove the lower bound for general graphs, instead of fixing b as the above did, they perform an adaptive thresholding where they adjust b according to U. For more details, please refer to their original paper.

Chapter 6

Generalization to sparse hypergraphs

6.1 Introduction

In previous chapters we reviewed some works done in the setting of graphs. The next natural challenge is to generalize these to hypergraphs. Recall that hypergraphs are generalization of graphs where each edge can contain possibly more than 2 vertices. We restrict our attention to r-uniform hypergraphs where each edge consists of exactly r vertices. Let $G^{(r)}_{n,p}$ be the r-uniform Erdős-Rényi random hypergraph on vertices $[n]$ where each r-element subset of $[n]$ is an edge with probability p independently. Given an r-uniform hypergraph H, the hypergraph homomorphisms $\mathrm{Hom}(H, G)$ and the homomorphism density function $t(H, G)$ could be defined analogously as

$$\mathrm{Hom}(H,G) = \{f : V(H) \to V(G) : \{f(v_1), ..., f(v_r)\} \in E(G) \text{ for every } \{v_1, ..., v_r\} \in E(H)\},$$

$$t(H,G) = \frac{|\mathrm{Hom}(H,G)|}{n^{v_H}}.$$

Let X_H be the random variable $t(H, G)$ where $G \sim G^{(r)}_{n,p}$. As in the graph case, we are interested in determining the asymptotics of the upper tail probability

$$\zeta^{(r)}(H,n,p,\delta) = \log \mathbb{P}\Big(X_H \geq (1+\delta)\mathbb{E}[X_H]\Big).$$

The set of weighted graphs $\mathcal{G}_n$ generalizes to $\mathcal{G}_n^{(r)}$, the set of functions from r element subsets of $[n]$ to $[0,1]$ (so that $\mathcal{G}_n = \mathcal{G}_n^{(2)}$) and the homomorphism counting function extends to $\mathcal{G}_n^{(r)}$ as

$$t(H,G) = n^{-v_H} \sum_{f:V(H)\to[n]} \prod_{\{v_1,...,v_r\}\in E(H)} G(f(v_1),...,f(v_r)).$$

Similar to (1.13), the variational problem ϕ could be defined analogously as

$$\phi^{(r)}(H,n,p,\delta) = \min_{G\in\mathcal{G}_n^{(r)}} \{I_p(G) : t(H,G) \geq (1+\delta)\mathbb{E}[t(H,G)]\}.$$

where $I_p(G) = \sum_{1\leq a_1<a_2<...<a_r\leq n} I_p(G(a_1,...,a_r))$.
Let Y_H be the random variable $N(H,G)$. Similar to (1.8) and (1.15), the upper tail for subgraph count R and the combinatorial optimization problem Φ are defined as

$$R^{(r)}(H,n,p,\delta) = \log \mathbb{P}(Y_H \geq (1+\delta)\mathbb{E}[Y_H]),$$

$$\Phi^{(r)}(H,n,p,\delta) = \min\{e_G \log(1/p) : G \subset K_n^{(r)} \text{ and } \mathbb{E}_G[Y_H] \geq (1+\delta)\mathbb{E}[Y_H]\}$$

where $\mathbb{E}_G[X_H] = \mathbb{E}[t(H,G')|G\subset G']$. When it is clear under context, we drop the "(r)" superscript from ζ, ϕ, R, Φ.
We would like to follow a similar two step strategy.

Step 1: Reduce ζ or R to the variational problem ϕ or the combinatorial optimization problem Φ.

Step 2: Solve the variational problem ϕ or combinatorial optimization problem Φ explicitly.

The work of [26] addresses step 2 for all hypergraph cliques $K_k^{(r)}$ for $r\geq 3$ and another special 3-uniform hypergraph as shown in Figure 6.1. Specifically, they proved the following.

Theorem 6.1.1 (Theorem 2.3 in [26])**.** *Fix an r-uniform hypergraph H and a real $\delta > 0$.*
(a) If $H = K_k^{(r)}$ for $r \geq 3$, then for any $n^{-1/\binom{k-1}{r-1}} \ll p \ll 1$ we have that

$$\lim_{n\to\infty} \frac{\phi(H,n,p,\delta)}{n^r p^{\binom{k-1}{r-1}} \log(1/p)} = \min\{\frac{\delta^{r/k}}{r!}, \frac{\delta}{(r-1)!k}\}.$$

(b) If H is the 3-graph in Figure 6.1, then for any $n^{-1/2} \ll p \ll 1$ we have that

$$\lim_{n\to\infty} \frac{\phi(H,n,p,\delta)}{n^3p^2\log(1/p)} = \frac{1}{6}\min\{\sqrt{9+3\delta}-3, \sqrt{\delta}\}.$$

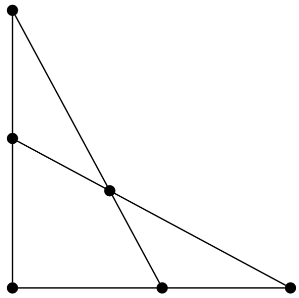

Figure 6.1 (Figure 1 in [26]) A 3-graph where dots denote vertices and each line denotes an edge.

For step 1, they apply a general nonlinear large deviations result from [16], thus obtaining an asymptotic for ζ in a narrower range of p. Instead we will be focusing on the first step, that is to reduce the upper tail probability to the combinatorial optimization problem Φ. We will achieve this for a special class of hypergraph cliques $K_{k+1}^{(k)}$. For simplicity of the proof, we will use the subgraph count function $N(H,G)$ instead of the homomorphism count function $t(H,G)$. We will prove Theorem 1.6.1, which we repeat here.

Theorem 6.1.2. *For fixed $k \geq 3$ and $p = p(n)$ such that $n^{-1}\log n \ll p \ll 1$,*

$$R(K_{k+1}^{(k)}, n, p, \delta) = -(1+o(1))\Phi(K_{k+1}^{(k)}, n, p, \delta + o(1)). \tag{6.1}$$

Note that Φ and ϕ behave the same asymptotically at least for the range specified in Theorem 6.1.1. To see this, for each $G \subset G_{k+1}^{(k)}$ we can construct $\mathcal{G} \in \mathcal{G}_{k+1}^{(k)}$ such that for each $e = (v_1, ...v_k)$,

$$\mathcal{G}(v_1,...,v_k) = \begin{cases} 1, & \text{for } e \notin G, \\ p, & \text{for } e \in G. \end{cases}$$

In the sparsity range of Theorem 6.1.1, the multiplicative difference between $t(H,G)$ and $N(H,G)$ does not matter, then $|\mathrm{Hom}(H,\mathcal{G})| \sim \mathbb{E}_G[N(H,G)]$ and $I_p(\mathcal{G}) = e_G \log(1/p)$. Thus $\Phi(K_{k+1}^{(k)}, n, p, \delta) \geq \phi(K_{k+1}^{(k)}, n, p, \delta)$. Together with the clique and hub construction from [26] which gives the same lower bound on Φ as on ϕ, we can conclude that ϕ and Φ

have the same asymptotic. Combining it with Theorem 6.1.2 gives us the following.

Corollary 6.1.3. *For $H = K_{k+1}^{(k)}$, if $n^{-1/k} \ll p \ll 1$, then*

$$\lim_{n\to\infty} \frac{\mathbb{P}\big(N(H,G) \geq (1+\delta)\mathbb{E}[N(H,G)]\big)}{n^k p^k \log(1/p)} = -\min\{\frac{\delta^{k/(k+1)}}{k!}, \frac{\delta}{(k-1)!(k+1)}\}.$$

In some sense, the range given in Theorem 6.1.1 is sub optimal since from what we learnt in the graph case, the asymptotic should hold as long as $\phi \to \infty$. Thus ignoring log factors, the range should be $n^{-1} \ll p$, which we achieve in Theorem 6.1.2. We note that any extension of the solution to the combinatorial optimization problem Φ to a wider range of p than in Theorem 6.1.1 would immediately yield a corresponding improvement in Corollary 6.1.3.

6.2 Notations

As in section 4.2, for a hypergraph H, let $\mathrm{Emb}(H,G) = \{f \in \mathrm{Hom}(H,G), f \text{ is injective}\}$ be the set of graph embeddings and $\mathrm{Emb}(H,G;e,e')$ be the set of embeddings that maps the edge e of H to the edge e' of G. For an r-uniform hypergraph G and $v_1, ..., v_{r-1} \in V(G)$, let $deg_G(v_1, ..., v_{r-1}) = |\{v : \{v, v_1, ..., v_{r-1}\} \in E(G)\}|$.

6.3 Proof of Theorem 6.1.2

It suffices to prove the following proposition.

Proposition 6.3.1. *Let $k \geq 3$ and $H = K_{k+1}^{(k)}$. For all positive $\delta > 0$ and $\epsilon < \min\{\delta, 1/2\}$, there exists a positive constant C such that the following holds. Suppose an integer n and $p \in (0,1)$ satisfy $Cn^{-1}\log n \leq p \leq 1/C$, then*

$$(1-\epsilon)\Phi(H,n,p,\delta-\epsilon) \leq -R(H,n,p,\delta) \leq (1+\epsilon)\Phi(H,n,p,\delta+\epsilon).$$

We will prove this with the entropic stability framework (in particular Theorem 4.1.3) established by [21]. The main idea is that every vertex of the core graph has to satisfy certain lower bound on the vertex degree, so we are able to construct a small set of induced hypergraphs from which the edges come from. This enables us to bound the number of cores.

First we introduce some results that lets us upper bound the number of sub-hypergraphs. Recall the fractional independence number α^* in Definition 4.4.3 which can be analogously defined for hypergraphs. The asymptotic order of $N(H,G)$ in terms of α^* was established in [18].

Theorem 6.3.2 (Theorem 1.1 in [18])**.** *For any hypergraph H, we have*

$$\max\{N(H,G) : e_G = l\} = \Theta(l^{\alpha^*_H}).$$

Lemma 6.3.3. *For $k \geq r$ and $H = K_k^{(r)}$, we have $\alpha^*_H = \frac{k}{r}$.*

Proof. Let $V(H) = [k]$ and consider the assignment α such that $\alpha(i) = \frac{1}{r}$ for every $i \in [k]$. This is a valid assignment so $\alpha^*_H \geq \frac{k}{r}$. On the other hand, for any valid assignment α and $e \in E(H)$, let $\alpha(e) = \sum_{v\in e} \alpha(v)$, then

$$e_H \geq \sum_{e\in E(H)} \alpha(e) = \sum_{v\in V(H)} \alpha(v) \cdot |\{e \in E(H) : v \in e\}| = \binom{k-1}{r-1} \sum_{v\in V(H)} \alpha(v).$$

Thus

$$\alpha^*_H \leq \frac{e_H}{\binom{k-1}{r-1}} = \frac{\binom{k}{r}}{\binom{k-1}{r-1}} = \frac{k}{r},$$

so we can conclude that $\alpha^*_H = \frac{k}{r}$. □

Combining Theorem 6.3.2 and Lemma 6.3.3 we have the following corollary.

Corollary 6.3.4. $\max\{N(K_{k+1}^{(k)}, G) : e_G = l\} \lesssim l^{\frac{k+1}{k}}$.

In the next lemma, we give some crude control on Φ in a range wider than Theorem 6.1.1 which is needed later in our proof.

Lemma 6.3.5. *Let $H = K_{k+1}^{(k)}$. For $n^{-1}\log n \leq p \ll 1$ we have*

$$(np)^{1/(k+1)} \log(1/p) \lesssim \Phi(H,n,p,\delta) \lesssim n^k p^k \log(1/p).$$

Proof. Suppose $G \subset K_n^{(k)}$ satisfies $\mathbb{E}_G[Y_H] \geq (1+\delta)\mathbb{E}[Y_H]$. By summing over all nonempty subgraph $J \subset H$ without isolated vertices, we have

$$\mathbb{E}_G[Y_H] - \mathbb{E}[Y_H] \leq \sum_{\emptyset\neq J\subset H} N(J,G) n^{k+1-v_J} p^{k+1-e_J} \leq 2^{k+1} \cdot \max_{\emptyset\neq J\subset H} e_G^{e_J} \cdot n^{k+1-v_J} p^{k+1-e_J}$$

$$\leq (2e_G)^{k+1} \cdot \frac{n^{k+1}p^{k+1}}{\min_{\emptyset \neq J \subset H} n^{v_J} p^{e_J}} \leq (2e_G)^{k+1} \cdot \frac{n^{k+1}p^{k+1}}{\min_{\emptyset \neq J \subset H} n^{v_J} p^{v_J}}$$

where the last inequality is due to the fact that for every $J \subset H$ we have $v_J \geq e_J$. Since $\mathbb{E}[Y_H] \asymp n^{k+1}p^{k+1}$ we have that $e_G \gtrsim (np)^{1/(k+1)}$, which gives us the lower bound on Φ. Next, consider a graph $G \subset K_n^{(k)}$ to be a clique of size $\lfloor (1+\delta)^{\frac{1}{k+1}} np \rfloor$ which contains at least $(1+\delta)\mathbb{E}[Y_H]$ copies of H. We then have

$$\Phi(H, n, p, \delta) \leq e_G \log(1/p) \lesssim n^k p^k \log(1/p).$$

$\square$

Now we are ready to prove our main proposition.

Proof of Proposition 6.3.1

Let G^* be a core graph and K be the constant given in Theorem 4.1.3. Recall that a core graph $G^* \in K_{k+1}^{(k)}$ satisfies

(C1) $\mathbb{E}_{G^*}[Y_H] \geq (1+\delta-\epsilon)\mathbb{E}[Y_H]$

(C2) $e_{G^*} \leq K \cdot \Phi(K_{k+1}^{(k)}, n, p, \delta+\epsilon)$

(C3) $\min_{e \in E(G^*)}(\mathbb{E}_{G^*}[Y_H] - \mathbb{E}_{G^* \setminus e}[Y_H]) \geq \mathbb{E}[Y_H]/(K \cdot \Phi(K_{k+1}^{(k)}, n, p, \delta+\epsilon))$

By Lemma 6.3.5 we can also choose the constant C from Proposition 6.3.1 large enough so that $\Phi(H, n, p, \delta-\epsilon) \gg K \log(1/p)$. By the entropic stability framework established in Theorem 4.1.3, the proposition would be implied by the following lemma that gives the upper bound on the number of core graphs G^*.

Lemma 6.3.6. *Assuming $Cn^{-1}\log n \leq p \leq 1/C$ for $C = C(\delta, \epsilon)$ sufficiently large, the number of cores with m edges is at most $(1/p)^{\epsilon m/2}$.*

Proof. From condition C3, for every edge $e \in E(G^*)$ we have

$$\mathbb{E}_{G^*}[Y_H] - \mathbb{E}_{G^* \setminus e}[Y_H] \leq \sum_{\emptyset \neq J \subset H} N(J, G^*; e) \cdot n^{k+1-v_J} p^{k+1-e_J},$$

where J ranges over all nonempty subgraph of H with no isolated vertices(recalling the notation $N(J, G; e)$ from Section 6.2). It follows that there is a subgraph $\emptyset \neq J \subset H$ and

$e' = \{v_1 ... v_k\} \in E(J)$ and constants $\gamma' = \gamma'(k)$ so that

$$\frac{|\mathrm{Emb}(J, G^*; e', e)|}{n^{v_J} p^{e_J}} \geq \frac{\gamma' \mathbb{E}[Y_H]}{K\Phi(H, n, p, \delta + \epsilon) n^{k+1} p^{k+1}}.$$

By Lemma 6.3.6, there is a constant $K' = K'(\delta)$ so that $\Phi(H, n, p, \delta+\epsilon) \leq K' n^k p^k \log(1/p)$. Combined with the fact that $\mathbb{E}[Y_H] \sim n^{k+1} p^{k+1}$, there is a constant $\gamma = \gamma(k)$ so that

$$\frac{Emb(J, G^*; e, e')}{n^{v_J} p^{e_J}} \geq \frac{\gamma}{KK' n^k p^k \log(1/p)}.$$

Thus

$$|\mathrm{Emb}(J, G^*; e, e')| \geq \frac{\gamma}{KK'} \cdot \frac{n^{v_J - k} p^{e_J - k}}{\log(1/p)}. \tag{6.2}$$

Since J is a nonempty subset of edges we have $v_J \geq k$. If $v_J = k$ we then have $e_J = 1$. Then

$$1 = |\mathrm{Emb}(J, G^*; e, e')| \geq \frac{\gamma}{KK'} \cdot \frac{(1/p)^{k-1}}{\log(1/p)}.$$

By letting C large enough and $p \to 0$ we get a contradiction. Thus $v_J = k+1$ and we get a similar contradiction if $e_J < k$ since fixing an edge of k vertices we only have one vertex left to embed and thus $|\mathrm{Emb}(J, G^*; e, e')| \leq n$. So J is either $J_1 := K_{k+1}^{(k)}$ or $J_2 := K_{k+1}^{(k)} \setminus e''$ for some edge e''. Without loss of generality, let $V(J) = \{v, v_1, ..., v_k\}$ and $e'' = \{v, v_2, ..., v_k\}$. Substituting J_1 and J_2 into (6.2) we get the following claim.

Claim 6.3.7. *For every edge $e \in E(G^*)$, either*

$$Emb(J_1, G^*; e', e) \geq x := \frac{\gamma}{KK'} \cdot \frac{np}{\log(1/p)}$$

or

$$Emb(J_2, G^*; e', e) \geq y := \frac{\gamma}{KK'} \cdot \frac{n}{\log(1/p)}.$$

Consider the following induced $k-1$ uniform hypergraphs $F_1 = F_1(G^*)$ and $F_2 = F_2(G^*)$ such that $V(F_1) = V(F_2) = \{1, ..., n\}$ and $\{i_1, ..., i_{k-1}\} \in E(F_1)$ if $\deg_{G^*}(i_1, ..., i_{k-1}) \geq x$ and $(i_1, ..., i_{k-1}) \in E(F_2)$ if $\deg_{G^*}(i_1, ..., i_{k-1}) \geq y$. Let $\eta = \frac{\gamma}{KK'}$, then we have

$$e(F_1) \leq \frac{km}{x} = \frac{km \log(1/p)}{\eta np}, e(F_2) \leq \frac{km}{y} = \frac{km \log(1/p)}{\eta n}.$$

Thus $\{v_1, ..., v_k\} \in E(G^*)$ implies that they either form a $K_k^{(k-1)}$ in F_1 or contain an edge in F_2. Letting $\mathcal{N}_m$ be the number of cores with m edges, we have

$$\mathcal{N}_m \lesssim \binom{\binom{n}{k-1}}{e_F}\binom{\binom{n}{k-1}}{e_H}\binom{N(K_k^{(k-1)},F_1)+e(F_2)\cdot n}{m}.$$

Now we bound the terms separately, we have

$$\log\binom{\binom{n}{k-1}}{e_{F_1}} \leq \log n^{(k-1)e_{F_1}} = \frac{(k-1)k\log n}{\eta np}\cdot m\log(1/p) \leq \frac{(k-1)k}{\eta C}\cdot m\log(1/p),$$

and similarly

$$\log\binom{\binom{n}{k-1}}{e_{F_2}} \leq n^{(k-1)e_{F_2}} = \frac{(k-1)k\log n}{\eta n}\cdot m\log(1/p) = o(m\log(1/p)).$$

By condition (C2) we have

$$m = e_{G^*} \leq KK'n^kp^k\log(1/p)$$

and by Corollary 6.3.4 we have $N(K_k^{(k-1)},F_1) \lesssim e_{F_1}^{\frac{k}{k-1}}$. Thus

$$\log\binom{N(K_k^{(k-1)},F_1)}{m} \lesssim m\log\frac{e_{F_1}^{\frac{k}{k-1}}}{m} \lesssim m\log\frac{m^{\frac{k}{k-1}}\log(1/p)^{\frac{k}{k-1}}}{mn^{\frac{k}{k-1}}p^{\frac{k}{k-1}}} = m\log\frac{m^{\frac{1}{k-1}}\log(1/p)^{\frac{k}{k-1}}}{n^{\frac{k}{k-1}}p^{\frac{k}{k-1}}}$$

$$\lesssim m\log\frac{n^{\frac{k}{k-1}}p^{\frac{k}{k-1}}\log(1/p)}{n^{\frac{k}{k-1}}p^{\frac{k}{k-1}}} = m\log\log(1/p) = o(m\log(1/p)),$$

and

$$\log\binom{e_{F_2}\cdot n}{m} \lesssim m\log\frac{e_{F_2}\cdot n}{m} \leq m\log\frac{\frac{km\log(1/p)}{\eta n}\cdot n}{m} \lesssim m\log\log(1/p) = o(m\log(1/p)).$$

Combining the above bounds we have that $\log\mathcal{N}_m \leq \frac{\epsilon}{2}m\log(1/p)$ for sufficiently large C and thus the lemma is proved. $\square$

Letting $\epsilon \to 0$ in Proposition 6.3.1 gives Theorem 6.1.2.

6.4 Speculations about general hypergraphs

Lower bounds for the number of restricted graph embeddings $|\mathrm{Emb}(J, G^*; e', e)|$ like (6.2) also hold for general hypergraphs. Thus the entropic stability property can be shown by giving an upper bound of $|\mathrm{Emb}(J, G^*; e', e)|$ in terms of the degree of the vertices of e. To achieve this in the graph case, [21] relies heavily on tools from graph theory, especially Lemma 4.4.4 on the decomposition of certain parts of the graph into disjoint edges and cycles. Lemma 4.4.4 implies that the fractional independence number α^* of a graph is a half-integer. For hypergraphs, we have the following theorem from [20].

Theorem 6.4.1 (Theorem 3.1 in [20])**.** *For any rational $p/q \geq 1$, there is a 3-uniform hypergraph H with $\tau^*(H) = p/q$ where τ^* stands for the fractional vertex covering number.*

The above result seems to suggest that Lemma 4.4.4 may not generalize easily to hypergraphs since the fractional independence number α^* of an r-uniform hypergraph may not be of the form $\frac{k}{r}$ for integer k. So in order to show entropic stability property for general hypergraphs, it may be necessary to develop a new set of hypergraph decomposition results.

References

[1] N. Alon. On the number of subgraphs of prescribed type of graphs with a given number of edges. *Israel J. Math.*, 38:116–130, 1981.

[2] A. Basak and R. Basu. Upper tail large deviations of regular subgraph counts in erdős-rényi graphs in the full localized regime. *Preprint, arXiv:1912.11410*, 2020.

[3] B. B. Bhattacharya, S. Ganguly, E. Lubetzky, and Y. Zhao. Upper tails and independence polynomials in random graphs. *Adv. Math.*, 319:313–347, 2017.

[4] B. Bollobás. *Random graphs: Cambridge University Press.* 2001.

[5] 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.*, 2007.

[6] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs ii. multiway cuts and statistical physics. *Ann. of Math.*, 2012.

[7] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Counting graph homomorphisms. In *Topics in discrete mathematics. Algorithms Combin.*, 2006.

[8] S. Chatterjee. The missing log in large deviations for triangle counts. *Random Structures Algorithms*, 40:437–451, 2010.

[9] S. Chatterjee. An introduction to large deviations for random graphs. *Bull. Amer. Math. Soc. (N.S.)*, 53(4):617–642, 2016.

[10] S. Chatterjee and A. Dembo. Nonlinear large deviations. *Adv. Math.*, 299:396–450, 2016.

[11] S. Chatterjee and S. R. S. Varadhan. The large deviation principle for the Erdős-Rényi Random Graph. *European J. Combin.*, 32:1000–1017, 2011.

[12] N. A. Cook and A. Dembo. Large deviations of subgraph counts for sparse Erdős-Rényi graphs. *Preprint, arXiv:1809.11148*, 2018.

[13] B. DeMarco and J. Kahn. Tight upper tail bounds for cliques. *Random Structures Algorithms*, 41:469–487, 2012.

[14] B. DeMarco and J. Kahn. Upper tails for triangles. *Random Structures Algorithms*, 40:452–459, 2012.

[15] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. 1998.

[16] R. Eldan. Gaussian-width gradient complexity, reverse log-Sobolev inequalities and narticlear large deviations. *Geom. Funct. Anal.*, 28:1548–1596, 2016.

[17] H. Finner. A generalization of holder's inequality and some probability inequalities. *Ann. Probab.*, 20:1893–1901, 1992.

[18] E. Friedgut and J. Kahn. On the number of copies of one hypergraph in another. *Israel J. Math.*, 105:251–256, 1998.

[19] A. M. Frieze and R. Kannan. A simple algorithm for constructing szemere'di's regularity partition. *Electron. J. Combin.*, 6, 1999.

[20] F. C. Graham, Z. Füredi, M. R. Garey, and R. L. Graham. On the fractional covering number of hypergraphs. *SIAM J. Discrete Math.*, 1:45–49, 1988.

[21] M. Harel, F. Mousset, and W. Samotij. Upper tails via high moments and entropic stability. *Preprint, arXiv:1904.08212*, 2019.

[22] S. Janson, K. Oleszkiewicz, and A. Rucinski. Upper tails for subgraph counts in random graphs. *Israel J. Math.*, 142:61–92, 2004.

[23] S. Janson and A. Rucinski. The infamous upper tail. *Random Structures Algorithms*, 20:317–342, 2002.

[24] S. Janson and A. Rucinski. The deletion method for upper tail estimates. *Combinatorica*, 24:615–640, 2004.

[25] J. H. Kim and V. H. Vu. Divide and conquer martingales and the number of triangles in a random graph. *Random Structures Algorithms*, 24:166–174, 2004.

[26] Y. P. Liu and Y. Zhao. On the upper tail problem for random hypergraphs. *Preprint, arXiv:1910.02916*, 2019.

[27] L. Lovász and B. Szegedy. Limits of dense graph sequences. *J. Combin. Theory Ser. B*, 96:933–957, 2004.

[28] L. Lovász. *Large Networks and Graph Limits.* 2012.

[29] E. Lubetzky and Y. Zhao. On the variational problem for upper tails in sparse random graphs. *Random Structures Algorithms*, 50:420–436, 2017.

[30] S. Mukherjee and B. B. Bhattacharya. A note on replica symmetry in upper tails of mean-field hypergraphs. *Preprint, arXiv:1812.09841*, 2018.

[31] B. Szegedy. Limits of kernel operators and the spectral regularity lemma. *European J. Combin.*, 32:1156–1167, 2011.

[32] S. R. S. Varadhan. Asymptotic probabilities and differential equations. *Comm. Pure Appl. Math.*, 19:261–286, 1966.

[33] V. H. Vu. A large deviation result on the number of small subgraphs of a random graph. *Comb. Probab. Comput.*, 10:79–94, 2001.