LARGE DEVIATIONS FOR RANDOM NETWORKS AND APPLICATIONS. LECTURE NOTES

SHIRSHENDU GANGULY

Abstract. While large deviations theory for sums and other linear functions of independent random variables is well developed and classical, the set of tools to analyze non-linear functions, such as polynomials, is limited. Canonical examples of such non-linear functions include subgraph counts and spectral observables in random networks.

In this notes, we review the recent exciting developments around building a suitable nonlinear large deviations theory to treat such random variables and understand geometric properties of large random networks conditioned on associated rare events.

We will start with a discussion on dense graphs and see how the theory of graphons provides a natural framework to study large deviations in this setting. We also discuss Exponential random graphs, a well known family of Gibbs measures on graphs, and the bearing this theory has on them. We will then review the new technology needed to treat sparse graphs. Finally, we will see how the above and new ideas can be used to study spectral properties in this context.

The lectures will aim to offer a glimpse of the different ideas and tools that come into play including from extremal graph theory, arithmetic combinatorics and spectral graph theory. Several open problems are also mentioned.

I will keep updating the notes, but please let me know if you spot any typos/errors or any references that I might have missed.

Contents

1. Introduction

Given a fixed graph H, the "infamous upper tail" problem, a name given by Janson and Rucinski [27], asks to estimate the probability that the number of copies of H in an Erdős-Rényi random graph exceeds its mean by some given constant factor. The following related problem was investigated by Chatterjee and Varadhan (2011).

Fix $0 < p < r < 1$ and consider an instance of an Erdős-Rényi random graph $G \sim G(n, p)$ with edge density p, conditioned to have at least as many triangles as the typical $G(n, r)$. Then is the graph G "close" to the random graph $G(n, r)$?

The lectures will review the recent exciting developments around the above and related questions (see also [12] for a wonderful account of the area).

2. Well known concentration inequalities and large deviations bounds.

We start by recalling a well known method to obtain standard concentration and large deviations bounds.

Proposition 2.1 (Azuma-Hoeffding). *Suppose $\{X_i\}$ is a martingale difference sequence with respect to some filtration $\mathcal{F}_i$. Also assume that almost surely given $\mathcal{F}_{i-1}$, $A_i \le X_i \le B_i$, where A_i, B_i are $\mathcal{F}_{i-1}$ measurable and $B_i - A_i \le c_i$ almost surely. Then for $S_n = \sum_{i=1}^n X_i$, for any $x > 0$,*

$$\mathbb{P}(|S_n| \ge x) \le 2\exp(-\frac{2x^2}{\sum_{i=1}^n c_i^2}).$$

Lemma 2.1. *(Hoeffding's lemma) Let X be a mean zero random variable such that $a \le X \le b$ almost surely, then for all $\theta \in \mathbb{R}$,*

$$\mathbb{E}(e^{\theta X}) \le e^{\theta^2(b-a)^2/8}.$$

Proof of Proposition 2.1. The proof follows by Cramer's method– Estimate exponential moment and apply Markov's inequality. To estimate $\phi_n(\theta) = \mathbb{E}(e^{\theta S_n})$, we first use the above lemma to notice that almost surely

$$\mathbb{E}(e^{\theta X_i}|\mathcal{F}_{i-1}) \le e^{\theta^2 c_i^2/8}.$$

Thus by induction,

$$\phi_n(\theta) \le e^{\frac{\theta^2}{8}\sum_{i=1}^n c_i^2}.$$

By Markov's inequality, for $x, \theta > 0$,

$$\mathbb{P}(S_n \ge x) \le e^{\frac{\theta^2}{8}\sum_{i=1}^n c_i^2 - \theta x}.$$

Optimize over θ to obtain

$$\mathbb{P}(S_n \ge x) \le e^{-\frac{2x^2}{\sum_{i=1}^n c_i^2}}.$$

The same bound can be obtained for $\mathbb{P}(S_n \le -x)$. ☐

Thus computing exponential moments gives us a natural way to obtain tail bounds. It turns out for special cases this provides optimal results.

2.1. **Coin tossing:** Let $X_1, X_2, \ldots$ be i.i.d. Bernoulli(p). Letting $S_n = \sum_{i=1}^n X_i$, of course, typically $S_n = np + O(\sqrt{n})$.

- What is $\mathbb{P}(S_n \ge nq)$ for some $q > p$?

Let $\Lambda(\theta) = \log(pe^\theta + 1 - p)$ be the log-moment generating function. Thus by the above strategy,

$$\log \mathbb{P}(S_n \ge nq) \le n(\lambda(\theta) - \theta q).$$

Relative entropy: $I_p(q) = q\log\frac{q}{p} + (1-q)\frac{(1-q)}{(1-p)}$ is the Legendre dual of $\Lambda(\theta)$, i.e.,

$$I_p(q) = \sup_\theta(\theta q - \Lambda(\theta)).$$

We hence get the finite sample "error free" bound

$$\log \mathbb{P}(S_n \ge nq) \le -nI_p(q). \tag{2.1}$$

2.2. **Lower bound (Tilting).** A general strategy is to come up with a measure under which the event $A = \{S_n \geq nq\}$ is typical and then estimate the change of measure cost.

Let $\mathbb{P} = \mathbb{P}_p$ be the product Bernoulli measure with density p and similarly define $\mathbb{P}_q$. Then

$$\mathbb{P}(A) = \int_A e^{\log \frac{d\mathbb{P}_p}{d\mathbb{P}_q}} d\mathbb{P}_q. \tag{2.2}$$

For any set of bits $\mathbf{x} = (x_1, x_2, \ldots, x_n)$,

$$\frac{d\mathbb{P}_p}{d\mathbb{P}_q}(\mathbf{x}) = \sum_i \left[x_i \log(\frac{p}{q}) + (1 - x_i) \log(\frac{1-p}{1-q}) \right].$$

By law of large numbers, under $\mathbb{P}_q$, this is typically $-nI_p(q)$. Moreover $\mathbb{P}_q(A) \approx 1/2$. Plugging this into (2.2) yields the lower bound

$$\log \mathbb{P}_p(A) \geq -nI_p(q) + o(n).$$

Linearity played a big role in the computation of the log-moment generating function.

3. Non-linear large deviations

Subgraph counts in random graphs- Let $\mathcal{G}_{n,p}$ be the Erdős–Rényi random graph on n vertices with edge probability p, and let X_H be the number of copies of a fixed graph H in it. The upper tail problem for X_H asks to estimate the large deviation rate function given by

$$-\log \mathbb{P}\left(X_H \geq (1+\delta)\mathbb{E}[X_H]\right) \quad \text{for fixed } \delta > 0\,.$$

Formally we will work with homomorphisms instead of isomorphisms, since unless p is very small, they agree up to smaller order terms. For any graph G with adjacency matrix $A = (a_{i,j})_{1 \leq i,j \leq n}$,

$$t(H, G) := n^{-|V(H)|} \sum_{1 \leq i_1, \cdots, i_k \leq n} \prod_{(x,y) \in E(H)} a_{i_x i_y}$$

and $X_H = t(H, \mathcal{G}_{n,p})$ is a polynomial of independent bits.

- How to bound: $P\left(X_H \geq (1+\delta)\mathbb{E}[X_H]\right)$?
- How does the graph look conditionally?

A guess is that it looks like an inhomogeneous random graph.

Although this was established first in the seminal paper of Chatterjee-Varadhan [15], we will present a similar, but slightly more combinatorial argument from [32].

Definition 3.1 (Discrete variational problem, cost of change of measure)**.** Let $\mathscr{G}_n$ denote the set of weighted undirected graphs on n vertices with edge weights in $[0,1]$, that is, if $A(G)$ is the adjacency matrix of G then

$$\mathscr{G}_n = \{G_n : A(G_n) = (a_{ij})_{1 \leq i,j \leq n},\ 0 \leq a_{ij} \leq 1,\ a_{ij} = a_{ji},\ a_{ii} = 0 \text{ for all } i,j\}\,.$$

Let H be a fixed graph of size k, with maximum degree Δ. The variational problem for $\delta > 0$ and $0 < p < 1$ is

$$\phi(H, n, p, 1+\delta) := \inf\left\{ I_p(G_n) : G_n \in \mathscr{G}_n \text{ with } t(H, G_n) \geq (1+\delta)p^{|E(H)|} \right\} \tag{3.1}$$

where

$$t(H, G_n) := n^{-|V(H)|} \sum_{1 \leq i_1, \cdots, i_k \leq n} \prod_{(x,y) \in E(H)} a_{i_x i_y}.$$

$I_p(G_n)$ is the entropy relative to p, that is,

$$I_p(G) := \sum_{1\le i<j\le n} I_p(a_{ij}) \quad \text{and} \quad I_p(x) := x\log\frac{x}{p} + (1-x)\log\frac{1-x}{1-p}\,.$$

A key tool to answer such questions is Szemerédi's regularity lemma (a fundamental result in extremal graph theory). To state it and its applications, it would be useful to define a metric on graphs and an embedding of all of them in the same space.

Definition 3.2. Let $\mathscr{W}$ be the set of all symmetric $f : [0,1]^2 \to [0,1]$, that are Borel measurable where we identify two elements that are equal almost surely, which will be called as *graphons*.

Sometimes one needs to consider functions which take values beyond $[0,1]$ such as in $[-1,1]$, or $\mathbb{R}$. Note that a graph naturally embeds into $\mathscr{W}$ as a measurable function which is $0-1$ valued.

Definition 3.3. The cut distance on $\mathscr{W}$ is defined by

$$d_\square(f,g) = \sup_{S,T}\left|\int\int_{S\times T} f(x,y) - g(x,y)dxdy\right|$$

Lemma 3.1 (Counting Lemma). *For any finite simple graph H and any $f,g \in \mathscr{W}$,*

$$|t(H,f) - t(H,g)| \le |E(H)| d_\square(f,g),$$

where

$$t(H,W) := \int_{[0,1]^{|V(H)|}} \prod_{(i,j)\in E(H)} W(x_i,x_j)\,\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_{|V(H)|}\,,$$

is the density of H in W.

Equivalence under isomorphisms. We want to identify two graphs/graphons if one is obtained from the other just by a relabeling of vertices. To this end

$$\delta_\square(f,g) = \inf_\sigma d_\square(f, g\circ\sigma),$$

where σ is a measure preserving bijection from $[0,1]$ to itself, and $g\circ\sigma(x,y) := g(\sigma(x),\sigma(y))$. We will denote the quotient space as $\widetilde{\mathscr{W}}$ and $\delta_\square$ as the induced metric on the same.

Definition 3.4 (Graphon variational problem). For $\delta > 0$ and $0 < p < 1$, let

$$\phi(H,p,1+\delta) := \inf\Big\{\tfrac{1}{2}I_p(W) : \text{graphon } W \text{ with } t(H,W) \ge (1+\delta)p^{|E(H)|}\Big\}, \tag{3.2}$$

where

$$I_p(W) := \int_{[0,1]^2} I_p(W(x,y))\,\mathrm{d}x\mathrm{d}y.$$

3.1. **Szemerédi's regularity lemma.** In essence it says that

"All graphs approximately look like stochastic block models where the number of blocks depends only on the error."

Let $G = (V,E)$ be a simple graph, and let X, Y be subsets of V . Let $E_G(X,Y)$ be the number of edges in G going from X to Y(edges whose endpoints belong to $X\cap Y$ are counted twice). Let

$$\rho_G(X,Y) = \frac{E_G(X,Y)}{|X||Y|}$$

be the edge density across X, Y. Given $\varepsilon > 0$, we say a pair of disjoint sets $A, B \subset V$, $\varepsilon-$regular if for all $X \subset A$ and $Y \subset B$, with $|X| \geq \varepsilon|A|, |Y| \geq \varepsilon|B|$,

$$|\rho_G(A, B) - \rho_G(X, Y)| \leq \varepsilon.$$

Theorem 3.1 (Frieze-Kannan Weak regularity lemma.)**.** *For every $k \geq 1$, every graph $G = (V, E)$ admits a partition $\mathcal{P}$ into k classes (say $A_1, A_2, \ldots, A_k$) such that*

$$\delta_{\square}(G, G_{\mathcal{P}}) \leq \frac{2}{\sqrt{\log k}},$$

where $G_{\mathcal{P}}$ is the weighted graph obtained by considering the complete graph on V, and putting weight $\rho_G(A_i, A_j)$ on every edge between A_i and A_j.

As a consequence of the above and the aforementioned counting lemma, the following holds.

Proposition 3.2. [32] *Given $\varepsilon > 0$, for any graph $G = (V, E)$ on n vertices, there exists a partition $\mathcal{P}$ of V into at most $4^{1/\varepsilon^2}$ parts, say $A_1, \ldots A_M$, such that if $\rho_{i,j} = \rho_G(A_i, A_j)$, then*

$$\Big|t(K_3, G) - \sum_{i,j,k} |A_i||A_j||A_k|\rho_{i,j}\rho_{j,k}\rho_{k,i}\Big| \lesssim \varepsilon n^3.$$

We will now try to bound the probability of different edge densities. Fixing a partition $A_1, A_2, \ldots A_M$ and densities $\{d_{i,j}\}_{1\leq i\leq j\leq m}$ (we will assume all the densities to be at least p), by the binomial probabilities large deviation bound (2.1)

$$\mathbb{P}(\rho_{i,j} \geq d_{i,j} \text{ for all } 1 \leq i \leq j \leq m) \leq \exp\left(\sum_{i<j} |A_i|A_j I_p(d_{i,j}) + \sum_i \binom{|A_i|}{2} I_p(d_{i,i})\right).$$

Thus if $\sum_{i,j,k} |A_i||A_j||A_k|d_{i,j}d_{j,k}d_{k,i} \geq (1+\delta)n^3p^3$, then the RHS above is at least $\phi(K_3, n, p, \delta)$. Choosing from an ε mesh of possible values of $d_{i,j}$ (ε^{-M^2} choices) and all possible parts (M^n choices), a union bound proves an upper bound of

$$\varepsilon^{-M^2} M^n e^{-\phi(K_3,n,p,1+\delta-\eta)}$$

where $\eta = O(\frac{\varepsilon}{p^3})$.

The pre-factors blow up if p goes to zero faster than a certain poly-log in n.

Chatterjee and Varadhan [15], proved a precise large deviation theory on $\widetilde{\mathscr{W}}$ and using the fact that subgraph counts are continuous, proved

$$\mathbb{P}\left(t(H, \mathcal{G}_{n,p}) \geq (1+\delta)p^{|E(H)|}\right) = \exp\left(-(1+o(1))\binom{n}{2}\phi(H, p, \delta)\right),$$

As a consequence, one obtains concentration around the set of minimizers.

What are the minimizers?

- Open Problem. Given H, For what values of p, δ, is the solution a constant function?

Lubetzky and Zhao [31] found the exact region for all regular graphs H, using an elegant application of a generalized version of Hölder's inequality. For triangles, it yields for any graphon W,

$$T(K_3, W) \leq \mathbb{E}(W^2)^{3/2},$$

which can be used to show that any $r \geq p$, such that $(r^2, I_p(r))$ is on the convex minorant of the function $x \to I_p(\sqrt{x})$, is in the replica symmetric region, which was shown to be tight by constructing a better candidate in the complement region.

Even the case of a path of length two is open!! In an unpublished work with Bhaswar Bhattacharya, we found an improved bound than what the straightforward application of Hölder's inequality yields, which only depends on the maximum degree.

Experiments predict that, in many settings, in the symmetry breaking region, the optimal solution should be a block model with two blocks (see e.g., [29, 28]).

4. Gibbs measures.

Gibbs variational principle: For any $f : \{0,1\}^n \to \mathbb{R}$, let $Z_n = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} e^{f(x)}$

$$\log(Z_n) = \sup_{\nu} [\mathbb{E}_\nu(f) - D(\nu||\mu)],$$

where the supremum is over all measures ν on the hypercube, and $D(\nu||\mu)$ is the relative entropy of ν with respect to μ (the uniform measure).

• **Exercise:** Prove it.

4.1. **Mean field approximations:** When can one just restrict to product measures to get a reasonable estimate.

• **Exercise:** Prove that when f is affine, the supremum is attained for a product measure. Explicitly describe the measure.

4.2. **Exponential Random graphs.** A particularly important subclass of such measures, capturing clustering properties, is obtained when the Hamiltonian is given by counts of subgraphs of interest, such as triangles. This is termed in the literature as the *Exponential Random Graph model* (ERGM). Thus more precisely, for $x \in \{0,1\}^{n(n-1)/2}$, where the configuration space is the set of all graphs on the vertex set $\{1, \cdots, n\}$, defining $N_G(x)$ be the number of labeled subgraphs G in x, given a vector $\underline{\beta} = (\beta_1, \ldots, \beta_s)$, the ERGM Gibbs measure is defined as

$$\pi(x) = \frac{1}{Z_n} \exp\Big(\sum_{i=1}^{s} \beta_i \frac{N_{G_i}(x)}{n^{|V(G_i)|-2}} \Big) = \frac{1}{Z_n} \exp(n^2 T), \tag{4.1}$$

$$\psi_n := \frac{1}{n^2} \log Z_n. \tag{4.2}$$

Usually G_1 is taken to be the number of edges, G_2 as the number of triangles, and so on.

Theorem 4.1 (Chatterjee-Diaconis [14])**.** *Using the large deviation theory on graphons, the following mean field behavior was established.*

(1)

$$\psi = \lim_{n\to\infty} \psi_n = \sup_{\tilde{h} \in \widetilde{W}} T(\tilde{h}) - I(\tilde{h}),$$

where $I(\tilde{h}) = \frac{1}{2} \int \int_{[0,1]^2} \big[h(x,y) \log h(x,y) + (1 - h(x,y)) \log(1 - h(x,y)) \big] \mathrm{d}x\mathrm{d}y$, *is up to a translation* $I_{1/2}(h)$

(2) If all the β_is are non-negative, then the extremal solutions are constants.

Replica-Symmetry: In particular, they proved that the ferromagnetic case $\beta_1, \cdots, \beta_s > 0$ falls in the replica symmetric regime i.e., the maximizers are given by the following constant functions, i.e.,

$$\lim_{n\to\infty} \frac{1}{n^2} \log Z_n(\underline{\beta}) = \sup_{u\in[0,1]} \Big(\sum_{i=1}^{s} \beta_i u^{|E_i|} - \frac{1}{2} I(u) \Big). \tag{4.3}$$

Informally, the above implies that if $u_1, \cdots, u_k \in [0,1]$ are the maximizers in (4.3), then the ERGM behaves like a mixture of the Erdős-Rényi graphs $G(n, u_i)$s, in an asymptotic sense. This was made precise in the more recent work by Eldan [18], and Eldan and Gross [19, 20].

To understand the solutions of (4.3), for $p > 0$, define two functions $\Psi_{\underline{\beta}}$ and $\varphi_{\underline{\beta}}$ by

$$\Psi_{\underline{\beta}}(p) := \sum_{i=1}^{s} 2\beta_i |E_i| p^{|E_i|-1}, \quad \varphi_{\underline{\beta}}(p) := \frac{\exp(\Psi_{\underline{\beta}}(p))}{1 + \exp(\Psi_{\underline{\beta}}(p))}. \tag{4.4}$$

It is easy to check that both the above functions are increasing in p. We say that $\underline{\beta}$ belongs to the *high temperature phase or is subcritical if $\varphi_{\underline{\beta}}(p) = p$ has a unique solution p^* which in addition satisfies $\varphi'_{\underline{\beta}}(p^*) < 1$. Whereas, $\underline{\beta}$ is said to be in the low temperature phase or is supercritical, if $\varphi_{\underline{\beta}}(p) = p$ has at least two solutions p^* with $\varphi'_{\underline{\beta}}(p^*) < 1$. When $\underline{\beta}$ is neither in the high nor low temperature phase, is called the critical temperature phase.* Note that by computing the gradient of the RHS in (4.3), all the maximizers of the same, satisfy

$$\varphi_{\underline{\beta}}(p) = p.$$

Thus in the subcritical phase, the mixture of Erdős-Rényi graphs $G(n, u_i)$ degenerates to a single Erdős-Rényi graph $G(n, p^*)$.

In another major advancement in this field, in [5], an alternate method of understanding ERGM through studying convergence to equilibrium for the GD was adopted. In particular, it was shown that GD is rapidly mixing, if $\underline{\beta}$ is subcritical and is slowly mixing if it is supercritical instead.

Even though the above results establish in a certain weak sense, that ERGM in the high temperature phase behaves like an Erdős-Rényi graph $G(n, p^*)$, several problems remain open, in particular pertaining to how this approximation can be made quantitative. We list below a few of them:

(1) Does the Glauber dynamics satisfy functional inequalities like Poincaré and Log-Sobolev inequalities?
(2) What kind of concentration of measure does the ERGM exhibit?
(3) Do natural observables like the number of edges in an ERGM satisfy a central limit theorem?

Answers to such questions have several potential applications including in testing of hypothesis problems in the statistical study of networks. Unfortunately, in spite of the above mentioned mean field behavior, contrary to classical spin systems and lattice gases, a detailed analysis of general ERGM has been out of reach so far. Thus, while a lot of refined results on fluctuation theory and concentration properties, have been established over the years, for the exactly solvable Curie-Weiss model, (Ising model on the complete graph), corresponding questions for the ERGM remain largely open.

We recently made some progress with Kyeongsik Nam in [21] for sub-criticial ERGM. To state the results, let $M = \binom{n}{2}$ and we will consider ERGM as a measure on $\mathcal{G}_n$ (the space of graphs on n vertices) naturally identified with $\{0,1\}^M$, with a random configuration being denoted by

$X = (X_1, X_2, \ldots, X_m)$. Now for an dimensional vector $\mathbf{v} = (v_1, \cdots, v_M)$ with $v_i \geq 0$, and $\|\mathbf{v}\|_2$ denoting the ℓ_2 norm, and $f : \mathcal{G}_n \mapsto \mathbb{R}$, we say that $f \in \mathrm{Lip}(\mathbf{v})$ if

$$|f(x) - f(y)| \leq \sum_{i=1}^{M} v_i \mathbf{1}\{x_i \neq y_i\}. \tag{4.5}$$

Theorem 4.2 (Ganguly, Nam [21])**.** *There exists a constant $c > 0$ such that for sufficiently large n, for any $f \in \mathrm{Lip}(\mathbf{v})$ and $t \geq 0$,*

$$\mathbb{P}(|f(X) - \mathbb{E}f(X)| > t) \leq e^{-ct^2/\|\mathbf{v}\|_2^2}. \tag{4.6}$$

Theorem 4.3 (Ganguly, Nam [21])**.** *For any sequence of positive integers m satisfying $m = o(n)$ and $m \to \infty$ as $n \to \infty$ we have the following. Consider any set of m different edges $i_1, \cdots, i_m \in E(K_n)$ that do not share a vertex. Then, the following central limit theorem for the normalized number of open edges among $i_1, \cdots, i_m$ holds:*

$$\frac{X_{i_1} + \cdots + X_{i_m} - \mathbb{E}[X_{i_1} + \cdots + X_{i_m}]}{\sqrt{\mathrm{Var}(X_{i_1} + \cdots + X_{i_m})}} \xrightarrow{d} N(0,1), \tag{4.7}$$

where $\xrightarrow{d}$ denotes weak convergence.

- Open problems:
  (1) Prove a full CLT for the number of edges in a sub-critical ERGM.
  (2) What are the analogous results in low temperature?

5. Solution to the variational problems when $p \to 0$.

Recall the discrete and continuous variational problems and the corresponding solutions $\phi(H, n, p, \delta)$ and $\phi(H, p, \delta)$.
- The constant function is never the solution.

To state the results we need some notation.

Definition 5.1 (Independence polynomial)**.** For a graph H, The *independence polynomial* of H is defined to be

$$P_H(x) := \sum_k i_H(k) x^k \,,$$

where $i_H(k)$ is the number of k-element independent sets in H (sets without any induced edges).

Definition 5.2 (Inducing on maximum degrees)**.** For a graph H with maximum degree Δ, let H^* be the induced subgraph of H on all vertices whose degree in H is Δ. (Note that $H^* = H$ if H is regular.)

Roots of independence polynomials were studied in various contexts (cf. [10, 11, 16] and their references); here, the unique positive x such that $P_{H^*}(x) = 1 + \delta$ will, perhaps surprisingly, give the leading order behavior of $\phi(H, n, p, 1 + \delta)$ (possibly capped at some maximum value if H happens to be regular).

Theorem 5.1 (Bhattacharya-Ganguly-Lubetzky-Zhao[8])**.** *Let H be a fixed connected graph with maximum degree $\Delta \geq 2$. For any fixed $\delta > 0$ and $n^{-1/\Delta} \ll p = o(1)$, the solution to the discrete variational problem (3.1) satisfies*

$$\lim_{n \to \infty} \frac{\phi(H, n, p, 1 + \delta)}{n^2 p^\Delta \log(1/p)} = \begin{cases} \min\left\{\theta \,,\, \frac{1}{2}\delta^{2/|V(H)|}\right\} & \textit{if } H \textit{ is regular,} \\ \theta & \textit{if } H \textit{ is irregular,} \end{cases}$$

where $\theta = \theta(H, \delta)$ is the unique positive solution to $P_{H^*}(\theta) = 1 + \delta$.

Let H be a graph with maximum degree $\Delta = \Delta(H)$; recall that H is *regular* (or Δ-*regular*) if all its vertices have degree Δ, and *irregular* otherwise. Starting with a weighted graph G_n with all edge-weights a_{ij} equal to p, we consider the following two ways of modifying G_n so it would satisfy the constraint $t(H, G_n) \geq (1 + \delta)p^{|E(H)|}$ of the variational problem (3.1).

(a) (Planting a clique) Set $a_{ij} = 1$ for all $1 \leq i, j \leq s$ for $s \sim \delta^{1/|V(H)|} p^{\Delta/2} n$. This construction is effective only when H is Δ-regular, in which case it gives $t(H, G_n) \sim (1 + \delta)p^{|E(H)|}$.

(b) (Planting an anti-clique) Set $a_{ij} = 1$ whenever $i \leq s$ or $j \leq s$ for $s \sim \theta p^{\Delta} n$ for $\theta = \theta(H, \delta) > 0$ such that $P_{H^*}(\theta) = 1 + \delta$, in which case $t(H, G_n) \sim (1 + \delta)p^{|E(H)|}$.

Thus the above result says that, for a connected graph H and $n^{-1/\Delta} \ll p \ll 1$, one of these constructions has $I_p(G_n)$ that is within a $(1+o(1))$-factor of the optimum achieved by the variational problem (3.1).

For example, when $H = K_3$, the clique construction has $I_p(G_n) \sim \frac{1}{2}s^2 I_p(1) \sim \frac{1}{2}\delta^{2/3} n^2 p^2 \log(1/p)$, while $P_{K_3}(x) = 1 + 3x$ so $\theta = \delta/3$ and the anti-clique construction has $I_p(G_n) \sim snI_p(1) \sim \frac{1}{3}\delta n^2 p^2 \log(1/p)$ (thus the clique wins if $\delta > 27/8$).

• This extends previous results [32, Theorems 1.1 and 4.1] from cliques to the case of a general graph H.

• The results extend (see [8]) to any *disconnected* graph H. The interplay between different connected components can then cause the upper tail to be dominated not by an exclusive appearance of either the clique or the anti-clique constructions (as was the case for any *connected* graph H, cf. Theorem 7.3), but rather by an interpolation of these.

• The assumption $p \gg n^{-1/\Delta}$ in Theorem 7.3 is essentially tight in the sense that the upper tail rate function undergoes a phase transition at that location [26]: it is of order $n^{2+o(1)}p^{\Delta}$ for $p \geq n^{-1/\Delta}$, and below that threshold it becomes a function (denoted $M_H^*(n,p)$ in [26]) depending on all subgraphs of H. In terms of the discrete variational problem (3.1), again this threshold marks a phase transition, as the anti-clique construction ceases to be viable for $p \ll n^{-1/\Delta}$ (recall that $s \sim \theta p^{\Delta} n$ in that construction). Still, as in [32, Theorems 1.1 and 4.1], our methods show that if H is *regular* and $n^{-2/\Delta} \ll p \ll n^{-1/\Delta}$, the solution to the variational problem is $(1 + o(1))\frac{1}{2}\delta^{2/|V(H)|}$ (i.e., governed by the clique construction).

5.0.1. *Why sparsity helps!!* Entropic variational problem can be reduced to polynomial variational problems with the aid of the following estimates derived in [32].

(1) If $0 \leq x \ll p$, then $I_p(p + x) \sim \frac{1}{2}x^2/p$, whereas when $p \ll x \leq 1 - p$ we have $I_p(p + x) \sim x \log(x/p)$.

(2) There is some constant $p_0 > 0$, such that for every $0 < p \leq p_0$,
$$I_p(p + x) \geq (x/b)^2 I_p(p + b) \qquad \text{for any } 0 \leq x \leq b \leq 1 - p - \log(1 - p)\,.$$

(3) There is some constant $p_0 > 0$ such that for every $0 < p \leq p_0$,
$$I_p(p + x) \geq x^2 I_p(1 - 1/\log(1/p)) \sim x^2 I_p(1) \qquad \text{for any } 0 \leq x \leq 1 - p\,.$$

5.0.2. *Range of p.* It is expected that the large deviation behavior undergoes a transition from the mean field behavior to Poisson like behavior as p becomes sparser (For a regular graph $H = (V_H, E_H)$, of degree Δ, the Poisson regime is given by $1 \ll np^{\Delta/2} \ll (\log n)^{\frac{1}{|V_H|-2}}$).

However the task of rigorously proving that is not complete, notwithstanding a sequence of exciting improvements [13, 18, 17, 1, 24, 2]. In particular, the recent beautiful work [24] settled this problem for cliques. Further, this work also pinned down the Poisson like behavior for all regular graphs in the above mentioned regime. Subsequently, very recently, in [2], following the broad approach [24], the case for all regular graphs has been settled as well.

The case of irregular graphs is more delicate. It was shown in [26] that the log-probability for $p \gg n^{-1/\Delta}$ is $n^2 p^\Delta$ (where Δ is the maximum degree), while below, the rate depends on subgraphs, and even the order of the log- probabilities is not well understood. For some counterexamples see [33].

5.1. **Arithmetic progressions: another related setting.** Let X_k denote the number of k-term arithmetic progressions (k-AP) in the random set Ω_p, where Ω is taken to be either $\mathbb{Z}/N\mathbb{Z}$ or $[N] := \{1, \dots, N\}$ throughout this paper, and Ω_p denotes the random subset of Ω where every element is included independently with probability p.

The work of Warnke [34] settled the question of the asymptotic order of $\mathbb{P}(X_k \geq (1+\delta)\mathbb{E}X_k)$ when $\Omega = [N]$. He showed that for fixed $\delta > 0$ and $k \geq 3$, there exists constants $c, C > 0$ (depending only on k) such that

$$p^{C\sqrt{\delta}Np^{k/2}} \leq \mathbb{P}(X_k \geq (1+\delta)\mathbb{E}X_k) \leq p^{c\sqrt{\delta}Np^{k/2}}, \tag{5.1}$$

as long as $p^{k/2} \gtrsim (\log N/N)$ and p is bounded away from 1, and the probability being $e^{-\Theta(N^2p^k)}$ for $\frac{1}{N} \lesssim p^{k/2} \lesssim (\log N/N)$. However, the natural question of precise asymptotics had remained open.

Theorem 5.2 (Bhattacharya-Ganguly-Shao-Zhao[9])**.** *Fix $k \geq 3$. Let $\Omega = [N]$. Let $p = p_N \to 0$ and $\delta > 0$ be fixed. Then, as $N \to \infty$,*

$$\phi_p^{(k,\Omega)}(\delta) = (1+o(1))\sqrt{\delta}p^{k/2}N\log(1/p)$$

where $\phi_p^{(k,\Omega)}(\delta)$ is the corresponding variational problem.

The lower bound can be seen by forcing an interval of length $(1+o(1))\sqrt{\delta}p^{k/2}N$ to be present in Ω, so that it generates the extra $\delta\mathbb{E}X_k$ many k-APs as desired. The above along with results of [18] settled the large deviations question for a certain sparsity range of p.

We prove Theorem 5.2 by first reducing the variational problem to an extremal problem in additive combinatorics, namely that of determining the size of the smallest subset of Ω with a given number of k-APs, or equivalently, the maximum number of k-APs in a subset of Ω of a given size.

Proposition 5.3. *Under the same hypotheses as Theorem 5.2, as $N \to \infty$,*

$$\phi_p^{(k,\Omega)}(\delta) = (1+o(1))\log(1/p) \cdot \min_{S \subset \Omega}\Big\{|S| : T_k(S) \geq \delta p^k T_k(\Omega)\Big\}. \tag{5.2}$$

The number of k-APs in a set of given (sufficiently small) size is always maximized by an interval, as stated precisely below. The theorem below was proved for 3-APs by Green and Sisask [22] and extended to all k-APs in [9].

Theorem 5.4. *Fix a positive integer $k \geq 3$. There exists some constant $c_k > 0$ such that the following statement holds. Let $A \subset \mathbb{Z}$ be a subset with $|A| = n$. Then $T_k(A) \leq T_k([n])$.*

Recently in [24], the optimal result has been achieved matching the regime indicated in (5.1).

6. Mean field variational principle

• Recall the Gibbs measure formulation. For $f : \{0,1\}^n \to \mathbb{R}$, let $Z_n = \mathbb{E}_p(e^{f(x)})$. The original approach was to compute such exponential moments followed by Markov's inequality. Another related approach is to think of f as being "approximately" 0 on a set A and $-\infty$ on its complement. In this case Z_n is approximately $\mathbb{P}_p(A)$.

The mean field approximation is exact when f is linear. One can try to form a general theory and come up with conditions under which the mean field solution is a reasonable approximation. This was developed in the breakthrough work of Chatterjee-Dembo [13] who proved that:

For a smooth f on $[0,1]^n$,

$$\log(Z_n) \approx \sup_{\nu}[\mathbb{E}_{\nu}(f) - I_p(\nu)],$$

where ν is a product measure as long as

- f has low gradient complexity: $\{\nabla f(x) : x \in \{0,1\}^n\}$, where ∇ denotes the gradient, has a small covering number.
- The ℓ_∞ norms of the partial second derivatives, $\|f_{i,j}\|_\infty$, are small.

The above quantifies approximately linear type functions for which the mean field approach gives an asymptotically sharp estimate estimate.

Eldan [18] improved the results by replacing the covering number constraint to a Gaussian mean width constraint and getting rid of the second double derivative criterion. Augeri [1] continued the program and using convex analysis obtained sharper results, in particular proving the large deviation principle for cycles for $\log^2(n)n^{-1/2} \ll p \ll 1$.

• Instead of covering the the space of gradients, one can also cover the space of all graphs directly as done in the dense case. This approach has been taken in Cook-Dembo [17] who constructed a spectral cover, while in [24], the authors took a purely combinatorial approach inspired by the classical moment argument in [25], classifying the graphs according to presence of structures called 'seeds' and 'cores' facilitating the large deviation event. The work [2] also proceeds by a refined understanding of such structures.

We will discuss the Cook-Dembo approach briefly, to show that the solution to the variational problem, $\phi(\cdot, n, p, \cdot)$ indeed governs the large deviation behavior.

Recall the entropy function $I_p(y)$ for any $y \in [0,1]^d$. For any convex set $\mathcal{K} \in \mathbb{R}^d$, let

$$I_p(\mathcal{K}) = \inf_{y \in \mathcal{K}} I_p(y).$$

The following lemma which is a consequence of convex duality is the starting point.

Lemma 6.1 ([17]). *For any closed convex set $\mathcal{K} \in \mathbb{R}^d$*

$$\mathbb{P}_p(\mathcal{K}) \leq e^{-I_p(\mathcal{K})}$$

To upper bound large deviation probabilities of a random variable X, one needs to cover the space by a small number of such convex sets such that on each of them X does not fluctuate too much.

Proposition 6.1 ([17]). *Let $h : [0,1]^n \to \mathbb{R}$. Suppose that there is a family of closed convex sets $\{\mathcal{B}_i\}_{i\in\mathcal{I}}$ in $\mathbb{R}^n$ such that it covers $\{0,1\}^n$ upto an exceptional set $\mathcal{E} \in \{0,1\}^n$, i.e.,*

$$\{0,1\}^n \subset \mathcal{E} \cup_{i\in\mathcal{I}} \mathcal{B}_i.$$

Furthermore, suppose there exists $\delta > 0$, such that for any $i \in \mathcal{I}$,

$$|h(y) - h(x)| \le \delta,$$

for all $x, y \in \mathcal{B}_i$. Then, for all t,

$$\mathbb{P}(X \ge t) \le |\mathcal{I}| e^{-\phi(h,n,p,t-\delta)} + \mathbb{P}(\mathcal{E})$$

where $\phi(h,n,p,t-\delta)$ is the solution to the discrete variational problem for the random variable h.

Proof is left as an exercise. We will next discuss the case when $h = t(K_3,\cdot)$.

Let Sym_n be the space of all $n \times n$ symmetric matrices. For any $A \in \mathrm{Sym}_n$, recall the Hilbert-Schmidt norm,

$$\|A\|_{\mathrm{HS}} = \big(\sum_{i,j} A_{i,j}^2\big)^{1/2} = \big(\sum_i \lambda_i^2\big)^{1/2},$$

where $|\lambda_1| \ge |\lambda_2| \ldots \ge |\lambda_n|$ are the eigenvalues of A arranged in decreasing order of absolute value. Also recall the operator norm $\|A\|_{\mathrm{op}} = |\lambda_1|$.

The approach in [17] is to cover the space of symmetric matrices by balls in the operator norm. Using the relation $|a^3 - b^3| \le 3|a-b|(a^2+b^2)$, it follows that for any two matrices X and Y,

$$|\operatorname{tr}(X^3) - \operatorname{tr}(Y^3)| \le 3\|X-Y\|_{\mathrm{op}}(\|X\|_{\mathrm{HS}}^2 + \|Y\|_{\mathrm{HS}}^2) \le 6n^2\|X-Y\|_{\mathrm{op}}.$$

To control the size of the net, observe for any adjacency matrix A,

$$\sum_i \lambda_i^2 \le n^2, \text{ and hence } |\lambda_k| \le \frac{n}{\sqrt{k}}.$$

Let $X_{\le k}$ be the projection of X on the top k eigenvalues and similarly define $X_{>k}$. Now note the following inequality where

$$\|X-Y\|_{\mathrm{op}} \le \|X_{>k}\|_{\mathrm{op}} + \|X_{\le k} - Y\|_{\mathrm{HS}}.$$

So it suffices to construct a Hilbert-Schmidt norm net on rank k matrices of norm at most n.

Lemma 6.2 ([17]). *There exists such an $n\delta - \mathrm{HS}$ net of size $\mathcal{N} \le \exp(O(kn\log(n/\delta)))$.*

6.1. **Arithmetic progressions.** For $k = 3$, [13] had used the well known connection between Fourier analysis and arithmetic progressions of length three to verify the low gradient complexity condition. For higher values of k, a random sampling method was used in [9], which along with results of [18] or [13] can be used to prove the large deviation principle for p polynomially small. As alluded to before, optimal results were obtained recently in [24], relying on Janson's inequality.

7. Spectral Large deviations

In this section we will see how some of the above techniques and new arguments can be used to analyze large deviation properties of spectral statistics for random graphs. Throughout this section for an adjacency matrix of size n, we will denote

$$\lambda_1 \geq \lambda_2 \geq \ldots, \geq \lambda_n$$

to be the eigenvalues in non-increasing order.

It is well known that for p not too sparse ($\gg \frac{1}{n}$ up to poly-log factors; will be stated precisely soon) $\lambda_1 \approx np$ with the leading eigenvector being close to the constant vector.

Theorem 7.1 (Cook and Dembo [17]). *For any $\theta > 1$ fixed and $n^{-\frac{1}{2}} \ll p \leq \frac{1}{2}$,*

$$-\log \mathbb{P}(\lambda_1(\mathcal{G}_{n,p}) \geq \theta np) = (1+o(1))\phi_1(n,p,(1+o(1))\theta),$$

where

$$\phi_1(n,p,\theta) := \inf \left\{ I_p(G_n) : G_n \in \mathscr{G}_n \text{ with } \lambda_1(G_n) \geq \theta np \right\}. \tag{7.1}$$

Theorem 7.2 (Bhattacharya-Ganguly [7]). *For any $\delta > 0$ fixed and $n^{-\frac{1}{2}} \ll p \ll 1$,*

$$\lim_{n\to\infty} \frac{-\log \mathbb{P}(\lambda_1(\mathcal{G}_{n,p}) \geq (1+\delta)np)}{n^2p^2\log(1/p)} = \min\left\{ \frac{(1+\delta)^2}{2}, \delta(1+\delta) \right\}. \tag{7.2}$$

The proof of the first result is by covering the set $\{\lambda_1(G) \geq \theta np\}$ by not too many closed convex sets of the following type:

$$A_{u,v} = \{A : u^\top Av \geq \theta np\},$$

where $u, v \in \mathbb{S}^{n-1}$. Union over a net for the vectors u, v form the desired cover.

The proof of the second result uses the earlier results for cycles. To this end, for $s \geq 1$, let C_s denote the cycle of length s.

Theorem 7.3 (Bhattacharya-Ganguly-Lubetzky-Zhao [8]). *For any $\theta > 1$ and $n^{-\frac{1}{2}} \ll p \ll 1$, the solution to the variational problem (3.1) satisfies*

$$\lim_{n\to\infty} \frac{\phi(C_s,n,p,\theta)}{n^2p^2\log(1/p)} = \min\left\{ \gamma\, , \tfrac{1}{2}(\theta-1)^{\frac{2}{s}} \right\},$$

where $\gamma = \gamma(C_s,t)$ is the unique positive solution to $P_{C_s}(\gamma) = t$, with $P_{C_s}(\cdot)$ is the independence polynomial of C_s.

Let $G_n \in \mathscr{G}_n$ be a weighted graph satisfying $\lambda_1(G_n) \geq (1+\delta)np$. We can now use the above result to obtain a lower bound on (7.1), using the following simple inequality: For $s \geq 1$ even,

$$t(C_s,G_n) = \frac{1}{n^s}\sum_{j=1}^{n} \lambda_j^s(G_n) \geq \left(\frac{\lambda_1(G_n)}{n}\right)^s \geq (1+\delta)^s p^s.$$

This implies,

$$\phi_1(n,p,1+\delta) \geq \phi(C_s,n,p,(1+\delta)^s).$$

Therefore, by Theorem 7.3, it follows that, for any s even,

$$\liminf_{n\to\infty} \frac{\phi_1(n,p,1+\delta)}{n^2p^2\log(1/p)} \geq \liminf_{n\to\infty} \frac{\phi(C_s,n,p,(1+\delta)^s)}{n^2p^2\log(1/p)} = \min\left\{ \bar{\gamma}, \frac{1}{2}\left((1+\delta)^s - 1\right)^{\frac{2}{s}} \right\}. \tag{7.3}$$

where $\bar{\gamma} = \bar{\gamma}(C_s,\delta)$ is the unique positive solution to $P_{C_s}(\bar{\gamma}) = (1+\delta)^s$, with $P_{C_s}(\cdot)$ as defined in (7.4).

Therefore, to complete the proof of the lower bound on (7.1) , it suffices to understand the asymptotics of (7.3), as $s \to \infty$. In turns out that the independence polynomial for cycles can be calculated in closed form in terms of Chebyshev polynomials, using the recursion[1]

$$P_{C_s}(x) = P_{C_{s-1}}(x) + xP_{C_{s-2}}(x)\,, \quad P_{C_2}(x) = 2x+1\,, \quad P_{C_3}(x) = 3x+1\,.$$

Then using the closely related recursion for Chebyshev polynomials gives,

$$P_{C_s}(x) = \frac{1}{2^{s-1}} \sum_{a=0}^{\lfloor \frac{s}{2} \rfloor} \binom{s}{2a} (1+4x)^a,$$

which simplifies to

$$P_{C_s}(x) = \left[\frac{1}{2}(\sqrt{1+4x}+1)\right]^s + \left[\frac{1}{2}(\sqrt{1+4x}-1)\right]^s. \tag{7.4}$$

Using this the equation $P_{C_s}(\bar{\gamma}) = (1+\delta)^s$ can be written as:

$$\left(1+\sqrt{1+4\bar{\gamma}}\right)^s \left[1+\left(\frac{\sqrt{1+4\bar{\gamma}}-1}{\sqrt{1+4\bar{\gamma}}+1}\right)^s\right] = (2(1+\delta))^s.$$

This shows $\eta := \lim_{s\to\infty} \bar{\gamma}(C_s, \delta)$ must satisfy the equation $1+\sqrt{1+4\eta} = 2(1+\delta)$, which solves to $\eta = \delta(1+\delta)$. Therefore, taking limit as $s \to \infty$, in (7.3) gives

$$\liminf_{n\to\infty} \frac{\phi_1(n,p,1+\delta)}{n^2p^2\log(1/p)} \geq \min\left\{\frac{(1+\delta)^2}{2}, \delta(1+\delta)\right\}, \tag{7.5}$$

using $\lim_{s\to\infty}((1+\delta)^s - 1)^{\frac{2}{s}} = (1+\delta)^2$. This completes the proof of the lower bound.

The proof of the upper bound proceeds by verifying that the minimum entropy configurations are attained by planting a *clique* or an *anti-clique* of the required size in the Erdős-Rényi graph.

7.1. **Second eigenvalue.** Next, we study the large deviations of the second largest eigenvalue of $\mathcal{G}_{n,p}$. To this end, we begin by considering the operator norm of the centered adjacency matrix, that is, $A(\mathcal{G}_{n,p}) - p\mathbf{1}\mathbf{1}'$, a problem of independent interest.[2] The following mean field approximation was proved in [17], analogous to Theorem 7.1.

Theorem 7.4 (Cook and Dembo [17])**.** *For* $\frac{\log n}{n} \lesssim p \leq \frac{1}{2}$ *and* $\theta \gg n^{\frac{1}{2}}$,

$$-\log \mathbb{P}(||A(\mathcal{G}_{n,p}) - p\mathbf{1}\mathbf{1}'||_{\mathrm{op}} \geq \theta) = (1+o(1))\phi_2(n,p,\theta+o(\theta)),$$

where

$$\phi_2(n,p,\theta) := \inf\left\{I_p(G_n) : G_n \in \mathscr{G}_n \textit{ with } ||A(G_n) - p\mathbf{1}\mathbf{1}'||_{\mathrm{op}} \geq \theta\right\}. \tag{7.6}$$

Remark 7.1. Note that the above result gives the upper tail asymptotics for $||A(\mathcal{G}_{n,p}) - p\mathbf{1}\mathbf{1}'||_{\mathrm{op}}$ for deviations growing faster than $n^{\frac{1}{2}}$. On the other hand, the typical value of $||A(\mathcal{G}_{n,p}) - p\mathbf{1}\mathbf{1}'||_{\mathrm{op}}$ is $(1+o(1))\sqrt{np}$.

[1]By the definition of the independence polynomial, for any graph H and vertex v in it, $P_H(x) = P_{H_1}(x) + xP_{H_2}(x)$, where H_1 is obtained from H by deleting v and H_2 is obtained from H by deleting v and all its neighbors.

[2]Recall that for a $n \times n$ symmetric matrix A, the operator norm is defined as: $||A||_{\mathrm{op}} = \sup_{||x||=1} ||Ax||$.

Recently, Guionnet and Husson [23] established a large deviations principle for the largest eigenvalue of n-dimensional Wigner matrices, rescaled by $n^{-\frac{1}{2}}$, whose independent, standardized entries have uniformly sub-Gaussian moment generating functions (which allow for Rademacher entries). Observe that even though $A(\mathcal{G}_{n,p}) - \mathbb{E}(A(\mathcal{G}_{n,p}))$ is a Wigner matrix, such uniform sub-Gaussian domination does not apply in the case when $p \ll 1$. Therefore, establishing the large deviations for $||A(\mathcal{G}_{n,p}) - p\mathbf{1}\mathbf{1}'||_{\text{op}}$ in the scale $\sqrt{np}$ in the sparse regime ($p \ll 1$), where the mean-field variational problem (7.6) no longer determines the rate of the upper tail, remains open.

As in Theorem 7.2, we can complement Theorem 7.4 by solving the variational problem $\phi_2(n, p, \theta)$, in the sparse regime, proving the precise upper tail asymptotics for $||A(\mathcal{G}_{n,p}) - p\mathbf{1}\mathbf{1}'||_{\text{op}}$, for deviations growing faster than $n^{\frac{1}{2}}$.

Theorem 7.5 (Bhattacharya-Ganguly []). *For $n^{-\frac{1}{2}} \ll p \ll 1$ and $\delta = \delta_n = O(1)$ such that $\delta np \gg n^{\frac{1}{2}}$,*

$$\phi_2(n, p, \delta np) = (1 + o(1))\tfrac{1}{2}\delta^2 n^2 p^2 \log(1/p). \tag{7.7}$$

This implies, under the same conditions on p and δ as above,

$$\lim_{n\to\infty} \frac{-\log \mathbb{P}(||A(\mathcal{G}_{n,p}) - p\mathbf{1}\mathbf{1}'||_{\text{op}} \geq \delta np)}{\frac{1}{2}\delta^2 n^2 p^2 \log(1/p)} = 1. \tag{7.8}$$

Corollary 7.6. *For $n^{-\frac{1}{2}} \ll p \ll 1$ and $\delta = \delta_n < 1$ such that $\delta np \gg n^{\frac{1}{2}}$,*

$$\lim_{n\to\infty} \frac{-\log \mathbb{P}(\lambda_2(\mathcal{G}_{n,p}) \geq \delta np)}{\frac{1}{2}\delta^2 n^2 p^2 \log(1/p)} = 1. \tag{7.9}$$

Observe that the above result is stated only for $\delta < 1$. The following remark gives a brief, high level discussion of the reason for such a restriction.

Remark 7.2. Note that typically $\lambda_1(\mathcal{G}_{n,p}) \approx np$ and the corresponding Frobenius eigenvector is 'approximately' the constant vector $\mathbf{1}$, while $\lambda_2(\mathcal{G}_{n,p}) \approx \sqrt{np}$. A key step involved in our analysis of the large deviation properties of the latter, is to project the matrix $A(\mathcal{G}_{n,p})$ on to the orthogonal complement of $\mathbf{1}$, and by Weyl's inequality reducing the large deviation problem of $\lambda_2(\mathcal{G}_{n,p})$ to a large deviation problem of the largest eigenvalue of the projected matrix, and thereafter relying on Theorem 7.5. However, such a reduction is tight only if $\mathbf{1}$ is an approximate Frobenius eigenvector even after forcing $\lambda_2(\mathcal{G}_{n,p}) \geq \delta np$. Not surprisingly, it turns out that the latter constraint is approximately the same as forcing $\lambda_2 \approx \delta np$, while trivially this also implies $\lambda_1(\mathcal{G}_{n,p}) \geq \delta np$. Thus, the condition $\delta < 1$ ensures that the atypical behavior for $\lambda_2(\mathcal{G}_{n,p})$, does not cause $\lambda_1(\mathcal{G}_{n,p})$ and its leading eigenvector to behave atypically (they are still approximately np and $\mathbf{1}$, respectively), making the above proof strategy yield tight results.

The corresponding problem for the lower tail is also interesting, which for the case of the largest eigenvalue is well understood. It follows from results in Lubetzky and Zhao [31, Proposition 3.9] for the dense case and, more generally, from Cook and Dembo [17, Theorem 1.21], that the lower tail problem for $\lambda_1(\mathcal{G}_{n,p})$ exhibits replica symmetry, that is, the corresponding variational problem is minimized by the constant function. More precisely, it is known that, for $\frac{\log n}{n} \ll p \leq \frac{1}{2}$ and $0 < q < p$ (such that $s := q/p \in (0, 1)$ is fixed),

$$\mathbb{P}(\lambda_1(\mathcal{G}_{n,p}) \leq q(n-1)) = e^{-(1+o(1))\binom{n}{2}I_p(q)}.$$

Remark 7.3. Since the mean-field approximation to the upper tail probability for $\lambda_1(\mathcal{G}_{n,p})$, as in Theorem 7.1, is expected to hold beyond $p \gg n^{-1/2}$, it makes sense to ask what happens to the variational problem $\phi_1(n,p,1+\delta)$, for $n^{-1} \ll p \ll n^{-\frac{1}{2}}$. However, unfortunately, the sparsity of the graph renders the cycle homomorphism density irrelevant, because, in this regime, it is determined by the smaller eigenvalues and is no longer related to the largeness of $\lambda_1(\mathcal{G}_{n,p})$. This is reflected in the fact that in the said sparsity regime, $t(C_s, \mathcal{G}_{n,p})$ is much larger than the number of injective homomorphisms of C_s into $\mathcal{G}_{n,p}$. It is worth pointing out that although not directly related to spectral information, current technology can be used to establish (see [8]), that for any $\theta > 1$ and $n^{-1} \ll p \ll n^{-\frac{1}{2}}$,

$$\tilde{\phi}(C_s, n, p, \theta) = (1+o(1))\tfrac{1}{2}(\theta-1)^{\frac{2}{s}} n^2 p^2 \log(1/p),$$

where $\tilde{\phi}$ is the analogous variational problem to (3.1) for the injective homomorphism.

However for extremely sparse graphs, notably including the case of constant average degree, one needs new ideas. The next section treats this based on the recent work of Bhaswar B. Bhattacharya, Sohom Bhattacharya, and myself [6].

7.2. **Sparser graphs.** The breakthrough result of Krivelevich and Sudakov [30, Theorem 1.1] shows that, for any $p \in (0,1)$, with high probability

$$\lambda_1(\mathcal{G}_{n,p}) = (1+o(1)) \max\left\{\sqrt{d_{(1)}(\mathcal{G}_{n,p})}, np\right\}, \tag{7.10}$$

where $d_{(1)}(\mathcal{G}_{n,p})$ is the maximum degree of the graph $\mathcal{G}_{n,p}$. This result has been substantially generalized to other edge eigenvalues and to inhomogeneous random graph models by Benaych-Georges, Bordenave, and Knowles [3, 4]. Using the asymptotics of the typical value of the maximum degree $d_{(1)}(\mathcal{G}_{n,p})$ (cf. [30, Lemma 2.2] and [4, Proposition 1.9]), it follows from (7.10) above that $\lambda_{(1)}(\mathcal{G}_{n,p})$ has a qualitative change in behavior in the threshold regime where np is comparable to $\sqrt{\log n / \log\log n}$. More precisely (cf. [4, Proposition 1.9], and [30, Theorem 1.1 and Lemma 2.1]),

$$\lambda_{(1)}(\mathcal{G}_{n,p}) = \begin{cases} (1+o(1))\sqrt{d_{(1)}(\mathcal{G}_{n,p})} & \text{if } 0 < p \ll \frac{1}{n}\sqrt{\frac{\log n}{\log\log n}}, \\ (1+o(1))np & \text{if } \frac{1}{n}\sqrt{\frac{\log n}{\log\log n}} \ll p < 1. \end{cases} \tag{7.11}$$

This shows that the typical value of the largest eigenvalue is governed by the maximum degree of the random graph below the threshold, whereas above the threshold it is determined by the total number of edges in the graph. In this section, we will study the upper and lower tail large deviations of the r largest eigenvalues of $\mathcal{G}_{n,p}$ below the threshold. More specifically, we will consider the following sparsity regime:

$$\log n \gg \log(1/np) \quad \text{and} \quad np \ll \sqrt{\frac{\log n}{\log\log n}}. \tag{7.12}$$

It can be shown that throughout this regime, for $r \geq 1$ fixed and $1 \leq a \leq r$,

$$\lambda_a(\mathcal{G}_{n,p}) = (1+o(1))\sqrt{L_p}, \text{ where } L_p = \frac{\log n}{\log\log n - \log(np)}. \tag{7.13}$$

This result is proved in [4, Proposition 1.9, Corollary 1.13] with the first condition in (7.12) replaced by $np \gtrsim 1$. The slightly weaker constraint $\log n \gg \log(1/np)$, which can be rewritten as $p \gg \frac{1}{n^{1+o(1)}}$, allows a wider range of sparsity. Moreover, the edge eigenvalues remain bounded with high probability whenever $\log n \lesssim \log(1/np)$, whereas they diverge for $\log n \gg \log(1/np)$ (see Lemma **??**). Hence, the regime in (7.12) covers the complete range of p where the typical values of the edge eigenvalues are unbounded and governed by the largest degrees.

7.2.1. **Joint Large Deviation of the Edge Eigenvalues**. For $r \geq 1$, define the *upper tail event for the top r eigenvalues* as

$$\mathrm{UT}_r(\boldsymbol{\delta}) = \{F \in \mathscr{G}_n : \lambda_1(F) \geq (1+\delta_1)\sqrt{L_p}, \ldots, \lambda_r(F) \geq (1+\delta_r)\sqrt{L_p}\}, \tag{7.14}$$

where $\boldsymbol{\delta} = (\delta_1, \delta_2, \cdots, \delta_r)$ for some $\delta_1 \geq \delta_2 \geq \ldots \geq \delta_r > 0$. The following theorem gives the precise asymptotics of the log-probability of the upper tail event in the regime of interest.

Theorem 7.7. *For p as in (7.12), $r \geq 1$ fixed, and $\delta_1 \geq \delta_2 \geq \ldots \geq \delta_r > 0$,*

$$\lim_{n\to\infty} \frac{-\log \mathbb{P}(\mathcal{G}_{n,p} \in \mathrm{UT}_r(\boldsymbol{\delta}))}{\log n} = \sum_{a=1}^{r}(2\delta_a + \delta_a^2), \tag{7.15}$$

where $\mathrm{UT}_r(\boldsymbol{\delta})$ is as defined above in (7.14).

Note that the above shows that the upper tail probabilities decay as polynomials in n, unlike for $1/\sqrt{n} \ll p \ll 1$, where we saw in the last section that for $r = 1$, the upper tail probabilities are of the form $e^{-(1+o(1))c(\delta_1)n^2p^2\log(1/p))}$, for some constant $c(\delta_1) > 0$, depending on δ_1.

Next, we consider the lower tail probability. To this end, for $0 < \delta_1 \leq \delta_2 \leq \cdots \leq \delta_r < 1$, denote the *lower tail event for the top r eigenvalues* as,

$$\mathrm{LT}_r(\boldsymbol{\delta}) = \{F \in \mathscr{G}_n : \lambda_1(F) \leq (1-\delta_1)\sqrt{L_p}, \ldots, \lambda_r(F) \leq (1-\delta_r)\sqrt{L_p}\}. \tag{7.16}$$

The lower tail asymptotics, which happens in log-log-scale, is given in the theorem below.

Theorem 7.8. *For p as in (7.12), $r \geq 1$ fixed, and $0 < \delta_1 \leq \delta_2 \leq \cdots \leq \delta_r < 1$,*

$$\lim_{n\to\infty} \frac{1}{\log n}\left(\log\log \frac{1}{\mathbb{P}(\mathcal{G}_{n,p} \in \mathrm{LT}_r(\boldsymbol{\delta}))}\right) = 2\delta_r - \delta_r^2, \tag{7.17}$$

where $\mathrm{LT}_r(\boldsymbol{\delta})$ is as defined above in (7.16).

Note that the results above show, while the upper tail probabilities decay as polynomials in n, the lower tail probabilities decay as stretch exponentially n. This is consistent with many other natural settings where upper tail events lead to localized effects, whereas lower tail events force a more global change and, hence, are typically much more unlikely. Perhaps, the most well known example of this is in the tail behavior of the Tracy-Widom distribution function $F_{TW}(\cdot)$, which arises as the scaling limit of the largest eigenvalue of a Gaussian Unitary Ensemble (GUE) matrix. Here, the logarithm of the right tail of the Tracy-Widom distribution function $-\log(1 - F_{TW}(x))$ decays as $x^{3/2}$, while the logarithm of the left tail $-\log F_{TW}(x)$ decays as x^3.

The proof for the upper tail relies on a novel structure theorem, obtained by building on estimates in [30], followed by an iterative cycle removal process, which shows, conditional on the upper tail large deviation event, with high probability the graph admits a decomposition in to a disjoint union of stars and a spectrally negligible part. On the other hand, the key ingredient in the proof of the lower tail is a Ramsey-type result which shows that if the K-th largest degree of a graph is not atypically small (for some large K depending on r), then either the top eigenvalue or the r-th largest eigenvalue is larger than that allowed by the lower tail event on the top r eigenvalues, thus forcing a contradiction. The above arguments reduce the problems to developing a large deviation theory for the extremal degrees which could be of independent interest.

7.2.2. **Joint Large Deviation of the Largest Degrees**. Given $\varepsilon_1 \geq \varepsilon_2 \geq \cdots \geq \varepsilon_r > 0$, denote the *upper tail event for the r largest degrees* by

$$\mathrm{degUT}_r(\boldsymbol{\varepsilon}) := \{F \in \mathscr{G}_n : d_{(1)}(F) \geq (1+\varepsilon_1)L_p, \ldots, d_{(r)}(F) \geq (1+\varepsilon_r)L_p\}. \tag{7.18}$$

Similarly, for $0 < \varepsilon_1 \leq \varepsilon_2 \leq \cdots \leq \varepsilon_r < 1$, denote the *lower tail event for the r largest degrees* by

$$\mathrm{degLT}_r(\boldsymbol{\varepsilon}) := \{F \in \mathscr{G}_n : d_{(1)}(F) \leq (1-\varepsilon_1)L_p, \ldots, d_{(r)}(F) \leq (1-\varepsilon_r)L_p\}. \tag{7.19}$$

The following two propositions state the asymptotics of the probabilities of the upper and lower tail events respectively.

Proposition 7.9. *For $\log n \gg \log(1/np)$ and $np \ll \log n$, $r \geq 1$ fixed, and $\varepsilon_1 \geq \ldots \geq \varepsilon_r > 0$,*

$$\lim_{n\to\infty} \frac{-\log \mathbb{P}(\mathcal{G}_{n,p} \in \mathrm{degUT}_r(\boldsymbol{\varepsilon}))}{\log n} = \sum_{a=1}^{r} \varepsilon_a. \tag{7.20}$$

Proposition 7.10. *For $\log n \gg \log(1/np)$ and $np \ll \log n$, $r \geq 1$ fixed, and $0 < \varepsilon_1 \leq \ldots \leq \varepsilon_r < 1$,*

$$\lim_{n\to\infty} \frac{1}{\log n} \left(\log\log \frac{1}{\mathbb{P}(\mathcal{G}_{n,p} \in \mathrm{degLT}_r(\boldsymbol{\varepsilon}))} \right) = \varepsilon_r, \tag{7.21}$$

where $\mathrm{degLT}_r(\boldsymbol{\varepsilon})$ is as defined in (7.19).

Remark 7.4. *Note that in both the above results we consider $np \ll \log n$, instead of $np \ll \sqrt{\log n / \log\log n}$ as in (7.12). This is because, even though the typical value of the largest eigenvalue undergoes a transition from $\sqrt{L_p}$ to np, when np is comparable to $\sqrt{\log n/\log\log n}$, the largest degrees continue to be determined by L_p as long as $np \ll \log n$ (see Lemma* **??** *below), and the results above give their joint large deviations probabilities in this entire regime.*

7.3. **Open problems.** Finally, the above works leaves many questions open. We list below a few of them.

(1) Extending Corollary 7.6 to the case $\delta > 1$. The key issue one faces is that $\lambda_2(\mathcal{G}_{n,p}) > np$ automatically guarantees that $\lambda_1(\mathcal{G}_{n,p}) > np$ as well, and in particular the Perron-Frobenius eigenvector will now be not close to the constant vector. Hence, one cannot automatically transfer the knowledge about the operator norm of $A(\mathcal{G}_{n,p}) - p\mathbf{1}\mathbf{1}'$ to the second eigenvalue of $A(\mathcal{G}_{n,p})$.
(2) Establishing a joint large deviation for cycle homomorphism densities of different sizes, and using these, or otherwise, obtain a joint LDP for $\lambda_1(\mathcal{G}_{n,p})$ and $\lambda_2(\mathcal{G}_{n,p})$, or of various spectral moments.
(3) Compute large deviation probabilities for the k-th largest eigenvalue $\lambda_k(\mathcal{G}_{n,p})$. Does the upper tail large deviation behavior of $\lambda_k(\mathcal{G}_{n,p})$ agree with the probability of planting k small cliques of appropriate sizes (up to negligible factors in the exponential scale) in some regime of the parameter space?

Acknowledgements

I would like to thank my coauthors Bhaswar Bhattacharya, Sohom Bhattacharya, Eyal Lubetzky, Kyeongsik Nam, Fernando Shao, and Yufei Zhao as well as Sourav Chatterjee and Amir Dembo for several discussions over the years.

References

[1] Fanny Augeri. Nonlinear large deviation bounds with applications to traces of wigner matrices and cycles counts in Erdős-Rényi graphs. *Annals of Probability, to appear*, 2020.

[2] Anirban Basak and Riddhipratim Basu. Upper tail large deviations of regular subgraph counts in Erdős-Rényi graphs in the full localized regime. *arXiv:1912.11410*, 2019.

[3] Florent Benaych-Georges, Charles Bordenave, and Antti Knowles. Spectral radii of sparse random matrices. *arXiv preprint arXiv:1704.02945*, 2017.

[4] Florent Benaych-Georges, Charles Bordenave, and Antti Knowles. Largest eigenvalues of sparse inhomogeneous erdős–rényi graphs. *Annals of Probability*, 47(3):1653–1676, 05 2019.

[5] Shankar Bhamidi, Guy Bresler, and Allan Sly. Mixing time of exponential random graphs. *Ann. Appl. Probab.*, 21(6):2146–2170, 2011.

[6] Bhaswar B Bhattacharya, Sohom Bhattacharya, and Shirshendu Ganguly. Spectral edge in sparse random graphs: Upper and lower tail large deviations. *arXiv preprint arXiv:2004.00611*, 2020.

[7] Bhaswar B. Bhattacharya and Shirshendu Ganguly. Upper tails for edge eigenvalues of random graphs. *SIAM Journal on Discrete Mathematics, to appear*, 2020.

[8] Bhaswar B. Bhattacharya, Shirshendu Ganguly, Eyal Lubetzky, and Yufei Zhao. Upper tails and independence polynomials in random graphs. *Advances in Mathematics*, 319(313–347), 2017.

[9] Bhaswar B. Bhattacharya, Shirshendu Ganguly, Xuancheng Shao, and Yufei Zhao. Upper tails for arithmetic progressions in a random set. *International Mathematics Research Notices*, to appear, 2018.

[10] J. I. Brown, C. A. Hickman, and R. J. Nowakowski. The independence fractal of a graph. *J. Combin. Theory Ser. B*, 87(2):209–230, 2003.

[11] J. I. Brown and R. J. Nowakowski. Average independence polynomials. *J. Combin. Theory Ser. B*, 93(2):313–318, 2005.

[12] Sourav Chatterjee. Large deviations for random graphs. *Lecture Notes in Math*, 2197, 2017.

[13] Sourav Chatterjee and Amir Dembo. Nonlinear large deviations. *Adv. Math.*, 299:396–450, 2016.

[14] Sourav Chatterjee and Persi Diaconis. Estimating and understanding exponential random graph models. *Ann. Statist.*, 41(5):2428–2461, 2013.

[15] Sourav Chatterjee and S. R. S. Varadhan. The large deviation principle for the Erdős-Rényi random graph. *European J. Combin.*, 32(7):1000–1017, 2011.

[16] Maria Chudnovsky and Paul Seymour. The roots of the independence polynomial of a clawfree graph. *J. Combin. Theory Ser. B*, 97(3):350–357, 2007.

[17] Nick Cook and Amir Dembo. Large deviations of subgraph counts for sparse Erdős-Rényi graphs. *arXiv:1809.11148*, 2018.

[18] Ronen Eldan. Gaussian-width gradient complexity, reverse log-Sobolev inequalities and nonlinear large deviations. *Geom. Funct. Anal.*, to appear, 2018.

[19] Ronen Eldan, Renan Gross, et al. Decomposition of mean-field gibbs distributions into product measures. *Electronic Journal of Probability*, 23, 2018.

[20] Ronen Eldan, Renan Gross, et al. Exponential random graphs behave like mixtures of stochastic block models. *The Annals of Applied Probability*, 28(6):3698–3735, 2018.

[21] Shirshendu Ganguly and Kyeongsik Nam. Sub-critical exponential random graphs: concentration of measure and some applications. *arXiv preprint arXiv:1909.11080*, 2019.

[22] Ben Green and Olof Sisask. On the maximal number of 3-term arithmetic progressions in subsets of $\mathbb{Z}/p\mathbb{Z}$. *Bull. Lond. Math. Soc.*, 40(6):945–955, 2008.

[23] Alice Guionnet and Jonathan Husson. Large deviations for the largest eigenvalue of rademacher matrices. *Annals of Probability, to appear*, 2020.

[24] Matan Harel, Frank Mousset, and Wojciech Samotij. Upper tails via high moments and entropic stability. *arXiv:1904.08212*, 2019.

[25] Svante Janson, Tomasz Łuczak, and Andrzej Rucinski. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

[26] Svante Janson, Krzysztof Oleszkiewicz, and Andrzej Ruciński. Upper tails for subgraph counts in random graphs. *Israel J. Math.*, 142:61–92, 2004.

[27] Svante Janson and Andrzej Ruciński. The infamous upper tail. *Random Structures Algorithms*, 20(3):317–342, 2002.

[28] Richard Kenyon, Charles Radin, Kui Ren, and Lorenzo Sadun. Multipodal structure and phase transitions in large constrained graphs. *Journal of Statistical Physics*, 168(2):233–258, 2017.

[29] Richard Kenyon, Charles Radin, Kui Ren, and Lorenzo Sadun. Bipodal structure in oversaturated random graphs. *International Mathematics Research Notices*, 2018(4):1009–1044, 2018.

[30] Michael Krivelevich and Benny Sudakov. The largest eigenvalue of sparse random graphs. *Combinatorics, Probability and Computing*, 12(1):61–72, 2003.

[31] Eyal Lubetzky and Yufei Zhao. On replica symmetry of large deviations in random graphs. *Random Structures Algorithms*, 47(1):109–146, 2015.

[32] Eyal Lubetzky and Yufei Zhao. On the variational problem for upper tails in sparse random graphs. *Random Structures Algorithms*, 50(3):420–436, 2017.

[33] Matas Šileikis and Lutz Warnke. A counterexample to the demarco-kahn upper tail conjecture. *Random Structures & Algorithms*, 55(4):775–794, 2019.

[34] Lutz Warnke. Upper tails for arithmetic progressions in random subsets. *Israel J. Math.* to appear.