MODERATE DEVIATIONS IN TRIANGLE COUNT

JOE NEEMAN, CHARLES RADIN, AND LORENZO SADUN

Abstract. We prove moderate deviations bounds for the number of triangles in a $\mathcal{G}(n,m)$ random graph, complementing recent results of Goldschmidt et al.

The moderate deviations of triangle density in $\mathcal{G}(n,m)$ graphs change qualitatively between the regime of the central limit theorem and the regime of large deviations, with those of Goldschmidt et al. extending the former and our results extending the latter; we also conjecture a precise form of sharp change between the regimes.

Our results can be interpreted as finite size effects in phase transitions in constrained edge-triangle graphs.

1. Introduction

We prove moderate deviations bounds for the number of triangles in a $\mathcal{G}(n,m)$ random graph, deviations between those of Goldschmidt et al. [1] and large deviations addressed earlier. For instance, with the notation that $\tau(G)$ is the triangle density of a $\mathcal{G}(n,m)$ graph G where $n \to \infty$ and $m = p\binom{n}{2} + O(1)$, for some $1/2 \le p < 1$ that is fixed as $n \to \infty$ and $n^{-3/4} \ll t \ll 1$, we prove that

$$\Pr\left(\tau(G) \le p^3 - t\right) = \exp\left(-\frac{\ln \frac{1-p}{p}}{2(1-2p)} t^{2/3} n^2 + o(t^2)\right). \tag{1}$$

The number of triangles in a random graph is a fundamental and surprisingly important random variable in the study of probabilistic combinatorics. The probabilistic behavior of these triangle counts is at least partially responsible for the development of many important methods related to concentration inequalities for dependent random variables, including Janson's inequality [2], the entropy method [3], martingale difference techniques in random graphs, and others [4].

The traditional point of view, as exemplified by the seminal paper by Janson and Rucínski [5], holds that the lower tail of the triangle count is easy to characterize while the upper tail is hard. This view stems at least partly from the fact that most earlier works studied the $\mathcal{G}(n,p)$ model, in which edges appear independently, each with probability p. In the $\mathcal{G}(n,m)$ model, in which the number of edges is fixed at m, the situation is rather more subtle. For example, one can easily see that under $\mathcal{G}(n,p)$, the number of triangles, $T(G)$, satisfies $\mathrm{Var}(T(G)) = \Theta(n^4)$, while under $\mathcal{G}(n,m)$, $\mathrm{Var}(T(G)) = \Theta(n^3)$. The distinction between the two models – especially in the lower tail – becomes even more pronounced at larger deviations. This can be intuitively explained by the fact that in $\mathcal{G}(n,p)$ one can easily "depress" the triangle count simply by reducing the number of edges: a graph G with edge number $|E(G)| \approx q\binom{n}{2}$ will typically have triangle density $\tau \approxeq q^3$, and the probability of seeing such

Date: January 19, 2021.

a graph under $\mathcal{G}(n,p)$ is of the order $\exp(-\Theta(n^2(p-q)^2))$; it follows that under $\mathcal{G}(n,p)$ we have

$$\Pr(\tau(G) \leq \mathbb{E}\tau(G) - t) \geq \exp(-\Omega(n^2 t^2)). \tag{2}$$

Under $\mathcal{G}(n,m)$, large deficits in the triangle density are much rarer than they are in $\mathcal{G}(n,p)$. At the scale of constant-order deficits, this was noticed in [6, 7], where it is proved that for $t = \Theta(1)$ and $\mathcal{G}(n,m)$ with $m = \lfloor \frac{1}{2}\binom{n}{2} \rfloor$,

$$\Pr(\tau(G) \leq \mathbb{E}\tau(G) - t) = \exp(-\Theta(n^2 t^{2/3})). \tag{3}$$

(They also found the exact leading-order term in the exponent.) Extensive numerical computations [6, 7] strongly suggest that the phenomenon is not limited to $m \approx \frac{1}{2}\binom{n}{2}$. At the other end of the scale, a recent result of Goldschmidt et al. [1] showed that for $n^{-3/2} \ll t \ll n^{-1}$ the lower triangle tail has a different behavior:

$$\Pr(\tau(G) \leq \mathbb{E}\tau(G) - t) = \exp(-\Theta(n^3 t^2)). \tag{4}$$

(Again, they also found the exact leading-order term in the exponent.) Since $t \leq \Theta(n^{-3/2})$ is within the range of the central limit theorem this leaves open the case of $n^{-1} \ll t \ll 1$. Noting that the two exponential rates (namely $n^2 t^{2/3}$ and $n^3 t^2$) cross over at $t = \Theta(n^{-3/4})$ it is natural to guess both

$$\Pr(\tau(G) \leq \mathbb{E}\tau(G) - t) = \begin{cases} \exp(-\Theta(n^3 t^2)) & \text{if } t \ll n^{-3/4} \\ \exp(-\Theta(n^2 t^{2/3})) & \text{if } n^{-3/4} \ll t \ll 1. \end{cases} \tag{5}$$

As our main result we prove the second of these two cases, and conjecture the first. We also prove some structural results on graphs with $\tau(G) \leq \mathbb{E}\tau(G) - \omega(n^{-3/4})$. These structural results provide a plausible explanation for the importance of $t = \Theta(n^{-3/4})$, namely that it is the threshold at which a single large negative eigenvalue of the adjacency matrix becomes responsible for almost all of the triangle deficit.

2. Context and references

For convenience we note some common asymptotics notation. We use $f = o(g)$ or $f \ll g$ to mean $\lim |f(n)|/g(n) = 0$, $f = O(g)$ to mean $\limsup f(n)/g(n) < \infty$, $f = \Omega(g)$ to mean $\liminf f(n)/g(n) > 0$, $f = \omega(g)$ or $f \gg g$ to mean $\lim |f|/g = \infty$, and $f = \Theta(g)$ to mean both $f = O(g)$ and $f = \Omega(g)$. The phrase "with high probability" means "with probability converging to 1 as $n \to \infty$," and we will also make use of probabilistic asymptotic notation: "$f = O(g)$ with high probability" means that for every $\epsilon > 0$ there exists $C > 0$ with $\limsup \Pr(f \geq Cg) \leq \epsilon$; "$f = o(g)$ with high probability" means that for every $\epsilon > 0$, $|f|/g \leq \epsilon$ with high probability; and analogously for Ω and ω.

We are studying the triangle density of $\mathcal{G}(n,m)$ graphs in the range $\tau(G) = p^3 - t$ for $n^{-3/4} \ll t \ll 1$. The case $0 \leq t \leq \Omega(n^{-3/2})$ is within the range of the central limit theorem (and it is covered by Janson's more general work on subgraph statistics [8]). The range $n^{-3/2} \ll t \ll n^{-1}$ is studied by [1]; they showed that in this regime,

$$\Pr(\tau(G) \leq p^3 - t) = \exp\left(-\frac{t^2 n^3}{2\sigma_p^2}(1 + o(1))\right), \tag{6}$$

where $\sigma_p^2 = \mathrm{Var}(\tau(G))/n^3$, which is of constant order. They also show an upper bound for larger t: for $n^{-3/2} \ll t \ll 1$,

$$\Pr(\tau(G) \leq p^3 - t) = \exp\left(-\Omega(t^2 n^3)\right). \tag{7}$$

We show that this is not tight for $n^{-3/4} \ll t \ll 1$. For example, we show that in this range,

$$\Pr(\tau(G) \leq p^3 - t) = \exp\left(-\Theta(t^{2/3} n^2)\right). \tag{8}$$

In the case $p \geq \frac{1}{2}$, we also derive more detailed results: we identify the leading constant in the exponent and prove some results on the spectrum of the adjacency matrix of G.

2.1. **Related work on random graphs.** Besides the work of [1], there is related work on large deviation principles (LDPs) for more general statistics, and LDPs for sparser graphs, notably in [9].

Moderate deviations in triangle count in $\mathcal{G}(n,m)$ can be seen from a different vantage based on [10]. That paper follows a series of works on the asymptotics of 'constrained' random graphs, in particular the asymptotics of $\mathcal{G}(n,m,t)$, graphs on n nodes constrained to have m edges and t triangles. A large deviation principle, using optimization over graphons, a variant of the seminal work [11] by Chatterjee and Varadhan, was used to prove various features of phase transitions between asymptotic 'phases', phases illustrated by the entropy-optimal graphons. (See also [12].) But in [10] numerical evidence showed that the transitions could be clearly seen in finite systems, using constrained graphs with as few as 30 vertices. From this perspective moderate deviations in triangle count can be understood as *finite size effects in a phase transition*. Asymptotically, entropy goes through a sharp ridge as the edge density/triangle density pair (ε, τ) passes through $(\varepsilon, \varepsilon^3)$ (Thms. 1.1,1.2 in [7]), and moderate deviations quantify how the sharp ridge rounds off at finite node number, somewhat as an ice cube freezing in water has rounded edges. The focus thus shifts to the infinite system, where emergent phases are meaningful, away from $\mathcal{G}(n,m,t)$ or $\mathcal{G}(n,m)$.

2.2. **Related work on random matrices.** Since we are studying the spectrum of the adjacency matrix, our methods mainly come from random matrix theory. Specifically, we are interested in large deviations of eigenvalues of the random adjacency matrices coming from our random graphs. The study of large deviations of eigenvalues is an active topic, but the results we aim for are somewhat atypical. Traditionally, "large deviations" refers to deviations on the order of the mean, so large deviations results for random matrices typically consider the event that the largest eigenvalue of a symmetric $n \times n$ matrix with i.i.d. mean-zero, variance-σ^2 entries is of order $\alpha\sqrt{n}$ for $\alpha > 2\sigma$; this is because the typical value of the largest eigenvalue is of order $2\sigma\sqrt{n}$. But because an eigenvalue of order n^β contributes $n^{3\beta}$ to the triangle count, and since we are interested in triangle deviation of orders $n^{9/4}$ through n^3, we will necessarily be interested in much larger eigenvalues.

Another difference in our work is that we consider several large eigenvalues simultaneously. This is because we need to consider the possibility that the triangle count is affected by several atypically large eigenvalues instead of just one.

In related work,

- Guionnet and Husson [13] showed an LDP for the largest eigenvalue for a family of random matrices that includes Rademacher matrices, which is essentially the case that we consider when $p = \frac{1}{2}$.
- Augeri [14] showed an LDP for the largest eigenvalue for random matrices whose entries have heavier-than-Gaussian tails.
- Battacharya and Ganguly [15] showed an LDP for the largest eigenvalue of an Erdős-Rényi graph. Their setting differs from the others in that the random matrices they consider are not centered (which makes a big difference when studying the largest eigenvalue).
- Augeri, Guionnet, and Husson [16] showed an LDP for the largest eigenvalue for most random matrices with subgaussian elements. This is essentially the same random matrices that we consider, with the main difference being that they are looking at eigenvalues of size $\Theta(\sqrt{n})$.

3. Triangle counts

Our general setting is: we let A be the adjacency matrix of a $\mathcal{G}(n, m)$ graph, where $n \to \infty$ and $m = p\binom{n}{2} + O(1)$, for some $p \in \mathbb{R}$ that is fixed as $n \to \infty$. We denote by $\tau(A)$ the triangle density of A.

Theorem 1. *If $\frac{1}{2} \le p < 1$ and $n^{-3/4} \ll t \ll 1$ then*

$$\Pr\left(\tau(A) \le p^3 - t\right) = \exp\left(-\frac{\ln\frac{1-p}{p}}{2(1-2p)} t^{2/3} n^2 + o(t^2)\right) \tag{9}$$

(with the convention that $\frac{\ln\frac{1-p}{p}}{1-2p} = 2$ when $p = \frac{1}{2}$). Moreover, conditioned on $\tau(A) \le p^3 - t$, with high probability we have

$$\lambda_n^3(A) = -tn^3(1 - o(1)) \tag{10}$$

and $\lambda_{n-1}^3(A) \ge -o(tn^3)$.

3.1. **Centering the matrix.** The main point of this section is that when considering the lower tail for triangle counts in $\mathcal{G}(n, m)$ graphs, it suffices to look at eigenvalues of the centered adjacency matrix. This might sound obvious, but there are two subtleties:

(1) It is important that we are looking at the lower tail, because the upper tail probabilities are controlled by perturbations to the largest eigenvector; this is exactly the eigenvector that gets destroyed when we center the adjacency matrix, so the eigenvalues of the centered adjacency matrix don't give much information about the upper tail probabilities.
(2) It is important that we are looking at $\mathcal{G}(n, m)$ and not $\mathcal{G}(n, p)$, because in $\mathcal{G}(n, p)$ the entropically favorable way to reduce the triangle count is to reduce the number of edges; again, this primarily affects the largest eigenvector and so is not related to the centered adjacency matrix.

Lemma 2. *Let A be the adjacency matrix of a graph with n vertices and m edges. For any $p \in \mathbb{R}$,*

$$\operatorname{tr}[(A - p\mathbf{1} + pI)^3] = \operatorname{tr}[A^3] - p^3n^3 + p^3n + 6mp(np - 2p + 1) + 3p^3n(n-1) - 3p\sum_i d_i^2, \tag{11}$$

where d_i is the degree of vertex i. Or, if $p = m/\binom{n}{2}$, then

$$\operatorname{tr}[(A - p\mathbf{1} + pI)^3] \le \operatorname{tr}[A^3] - p^3n^3 + p^3n + 6mp. \tag{12}$$

Note that if A is sampled from $\mathcal{G}(m,n)$ and $p = m/\binom{n}{2}$ then $\mathbb{E}A = p\mathbf{1} - pI$, and so the quantity of interest in Lemma 2 is in fact the centered adjacency matrix $A - \mathbb{E}A$.

Proof. We expand everything in gory detail:

$$\begin{aligned}
&\operatorname{tr}[(A - p\mathbf{1} + pI)^3] - \operatorname{tr}[A^3] + p^3\operatorname{tr}[\mathbf{1}^3] - p^3\operatorname{tr}[I^3] \\
&= 3\operatorname{tr}[-pA^2\mathbf{1} + p^2A\mathbf{1}^2 + pA^2I + p^2AI^2 + p^3\mathbf{1}^2I - p^3\mathbf{1}I^2 - 2p^2A\mathbf{1}I] \\
&= 3\operatorname{tr}[-pA^2\mathbf{1} + np^2A\mathbf{1} + pA^2 + p^2A + np^3\mathbf{1} - p^3\mathbf{1} - 2p^2A\mathbf{1}] \\
&= 3(-p\operatorname{tr}[A^2\mathbf{1}] + 2(n-2)p^2m + 2pm + 0 + n^2p^3 - np^3),
\end{aligned}$$

where we used the fact that $\operatorname{tr}[A\mathbf{1}]$ is the sum of the entries of A and $\operatorname{tr}[A^2]$ is the sum of squares of entries of A; since A is an adjacency matrix, both of these are $2m$. Finally,

$$\operatorname{tr}[A^2\mathbf{1}] = \sum_{i,k}(A^2)_{ik} = \sum_{i,j,k} A_{ij}A_{jk} = \sum_j d_j^2. \tag{13}$$

This proves the equality. For the inequality, Cauchy-Schwarz implies that $\sum_i d_i^2 \ge \frac{1}{n}(\sum_i d_i)^2 = \frac{4m^2}{n} = 2mp(n-1)$. After applying this inequality and rewriting $3p^3n(n-1)$ as $6p^2m$, we obtain the inequality. □

Combining Lemma 2 with the observation that, for A the adjacency matrix of a $\mathcal{G}(n,m)$ graph, $\mathbb{E}\operatorname{tr}[A^3] = p^3n^3 + O(n^2)$, we arrive at the following consequence:

Corollary 3. *Let A be the adjacency matrix of a $\mathcal{G}(n,m)$ graph and let $\tilde{A} = A - \mathbb{E}A$. For any $t \ge 0$,*

$$\Pr(\operatorname{tr}[A^3] \le \mathbb{E}\operatorname{tr}[A^3] - t) \le \Pr(\operatorname{tr}[\tilde{A}^3] \le -t + O(n^2)) \tag{14}$$

For an inequality in the other direction, note that by the same argument as in Lemma 2, as long as $\sum_i d_i^2 \le n^3p^2 + D$, we have

$$\operatorname{tr}[(A - p\mathbf{1} + pI)^3] = \operatorname{tr}[A^3] - p^3n^3 + O(D + n^2). \tag{15}$$

Corollary 4. *With the notation of Corollary 3, if $D = \Omega(n^2)$ then*

$$\Pr\left(\operatorname{tr}[A^3] \le \mathbb{E}\operatorname{tr}[A^3] - t\right) \ge \Pr\left(\operatorname{tr}[\tilde{A}^3] \le -t + O(D) \text{ and } \sum_i d_i^2 \le n^3p^2 + D\right). \tag{16}$$

4. Large deviations for eigenvalues of random matrices

In this section and beyond, we will let A denote a generic random matrix. The most important such matrix is the centered adjacency matrix, previously denoted $\tilde{A}$ or $A - \mathbb{E}A$.

Definition 5. *For a random variable ξ, its cumulant-generating function is*

$$\Lambda_\xi(s) = \ln \mathbb{E} \exp(s\xi) \tag{17}$$

whenever the expectation exists; when the expectation does not exist, we set $\Lambda_\xi(s) = +\infty$.

Definition 6. *The random variable ξ is* subgaussian *if there exists a constant C such that $\Lambda_\xi(t) \le Ct^2$ for every $t \in \mathbb{R}$.*

Note that according to our definition, a subgaussian random variable has mean zero (since if $\Lambda_\xi(t)$ is finite on a neighborhood of 0 then $\Lambda_\xi(0) = 0$ and $\Lambda_\xi'(0) = \mathbb{E}\xi$, and so if $\mathbb{E}\xi$ is non-zero then one cannot have $\Lambda_\xi(t) \le Ct^2$ on a neighborhood of 0). Note also that if $\mathbb{E}\xi = 0$ and $\|\xi\|_\infty < \infty$ then ξ is subgaussian.

Definition 7. *For a function $f : \mathbb{R} \to \mathbb{R}$, its Legendre transform is the function $f^* : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ defined by*

$$f^*(y) = \sup_{x \in \mathbb{R}} \{xy - f(x)\} \tag{18}$$

Some basic properties of the Legendre transform include:

- If $f \le g$ then $f^* \ge g^*$.
- If f is convex then $f^{**} = f$.
- If $f(x) = cx^2$ then $f^*(x) = \frac{x^2}{4c}$.

Our goal in this note is to establish large deviations principles for extreme eigenvalues and singular values of random matrices. We will consider a symmetric $n \times n$ random matrix A_n (or sometimes just A) having i.i.d. upper-diagonal entries and zero diagonal entries. The letter ξ will always denote a random variable that is distributed as an upper-diagonal element of A, and we will always assume that ξ is subgaussian. We write $\lambda_i(A)$ for the eigenvalues of A (in non-increasing order) and $\sigma_i(A)$ for the singular values of A (in non-increasing order).

Theorem 8. *Let ξ be a subgaussian random variable. For any integer $k \ge 1$ and any sequence m_n satisfying $\sqrt{n} \ll m_n \ll n$, the sequence*

$$\frac{1}{m_n}(\sigma_1(A_n), \dots, \sigma_k(A_n)) \tag{19}$$

satisfies an LDP with speed m_n^2 and good rate function $I : \mathbb{R}_+^k \to [0, \infty)$ given by

$$I(x) = \frac{|x|^2}{2} \inf_{s \in \mathbb{R}} \frac{\Lambda^*(s)}{s^2}. \tag{20}$$

If we assume in addition that the function $s \mapsto \frac{\Lambda^*(s)}{s^2}$ achieves its infimum at some $s \ge 0$, then the sequence

$$\frac{1}{m_n}(\lambda_1(A_n), \dots, \lambda_k(A_n)) \tag{21}$$

satisfies an LDP with speed m_n^2 and the same good rate function I as above.

In the case that $\frac{\Lambda^*(s)}{s^2}$ saturates its infimum only at negative s, we are not able to show an LDP for the eigenvalues. Note, however, that $\sum_i \sigma_i^2(A) \geq \sum_i \lambda_i^2(A)$ and so our LDP for singular values provides an upper bound: it implies, for example, that

$$\frac{1}{m_n^2} \ln \Pr\left(\sqrt{\sum_i \lambda_i^2(A_n)} > m_n t\right) \leq -\frac{t^2}{2} \inf_{s \in \mathbb{R}} \frac{\Lambda^*(s)}{s^2} + o(1) \tag{22}$$

On the other hand, we can also easily show the lower bound

$$\frac{1}{m_n^2} \ln \Pr\left(\sqrt{\sum_i \lambda_i^2(A_n)} > m_n t\right) \geq -\frac{t^2}{2} \inf_{s \geq 0} \frac{\Lambda^*(s)}{s^2} - o(1), \tag{23}$$

but the assumption that $\frac{\Lambda^*(s)}{s^2}$ saturates its infimum only at negative s implies that these bounds are non-matching.

There are natural examples (including the most relevant one to us, the case of Bernoulli random variables with $p \geq \frac{1}{2}$), where $s^{-2}\Lambda^*(s)$ is increasing for $s \geq 0$. In this case,

$$\inf_{s \geq 0} s^{-2}\Lambda^*(s) = \lim_{s \to 0} s^{-2}\Lambda^*(s) = \frac{1}{2}(\Lambda^*)''(0) = \frac{1}{2\mathbb{E}\xi^2}, \tag{24}$$

and so our lower bound (for simplicity, focussing only on the case $k = 1$) becomes

$$\frac{1}{m_n^2} \ln \Pr\left(\lambda_1(A_n) > m_n t\right) \geq -\frac{t^2}{4\mathbb{E}\xi^2} - o(1). \tag{25}$$

When ξ has a Gaussian distribution, this turns out to be sharp, but we show that it is not sharp in general.

Theorem 9. *In the setting of Theorem 8, if $\mathbb{E}\xi^3 < 0$ and $\lim_{s \to \infty} s^{-2}\Lambda(s) = 0$ then there exists some $\eta > 0$ such that for any $t > 0$,*

$$-\lim_{n \to \infty} \frac{1}{m_n^2} \ln \Pr\left(\lambda_1(A_n) > m_n t\right) > -(1 - \eta)\frac{t^2}{4\mathbb{E}\xi^2}. \tag{26}$$

In particular, the assumptions of Theorem 9 are satisfied for the (centered) Bernoulli random variable with $p > \frac{1}{2}$.

For our applications to random graphs, we require a version of Theorem 8 for random bits chosen without replacement. Specifically, we consider the Erdős-Rényi random graphs $\mathcal{G}(n, m)$, where m is an integer satisfying $|m - p\binom{n}{2}| = O(1)$ (and $p \in (0, 1)$ is fixed). If A_n is the centered adjacency matrix of $\mathcal{G}(n, p)$ then it is covered by Theorem 8, where ξ is the random variable taking the values $-p$ and $1 - p$ with probabilities $1 - p$ and p respectively. In this case, we have

$$\Lambda_\xi^*(s) = D(p + s, p) := (p + s) \ln \frac{p + s}{p} + (1 - p - s) \ln \frac{1 - p - s}{1 - p}, \tag{27}$$

with the understanding that $\Lambda_\xi^*(s) = +\infty$ whenever $p + s \notin (0, 1)$. It is not hard to check that $\frac{\Lambda_\xi^*(s)}{s^2}$ achieves its infimum at some $s \geq 0$ if and only if $p \leq \frac{1}{2}$. Moreover, the conclusions of Theorem 8 remain true for the $\mathcal{G}(n, m)$ model.

Theorem 10. *Fix $p \in (0,1)$ and let A_n be the centered adjacency matrix of a $\mathcal{G}(n,m)$ random graph. For any integer $k \geq 1$ and any sequence m_n satisfying $\sqrt{n} \ll m_n \ll n$, the sequence*

$$\frac{1}{m_n}(\sigma_1(A_n), \dots, \sigma_k(A_n)) \tag{28}$$

satisfies an LDP with speed m_n^2 and good rate function $I : \mathbb{R}_+^k \to [0,\infty)$ given by $I(x) = \frac{|x|^2}{2} \cdot \frac{\ln \frac{1-p}{p}}{1-2p}$ (or $I(x) = |x|^2$ when $p = \frac{1}{2}$).

If, in addition, $p \leq \frac{1}{2}$ then the sequence

$$\frac{1}{m_n}(\lambda_1(A_n), \dots, \lambda_k(A_n)) \tag{29}$$

also satisfies an LDP with the same speed and rate function.

5. Some preliminaries

Most of this work is concerned with handling the triangle-count contribution of very negative eigenvalues, but we also need to show that there is no significant contribution from the rest. For this, we will use a deviation inequality from [17]:

Theorem 11. *Assume that $\|\xi\|_\infty < \infty$, and let $f : \mathbb{R} \to \mathbb{R}$ be a 1-Lipschitz, convex function. Define $X_n = \frac{1}{n}\sum_{i=1}^n f(n^{-1/2}\lambda_i(A_n))$. Then there is a universal constant $C < \infty$ such that for any $\delta \gg n^{-1}$,*

$$\Pr(|X_n - \mathbb{E}X_n| \geq \delta) \leq C \exp\left(-\frac{n^2\delta^2}{C\|\xi\|_\infty^2}\right). \tag{30}$$

Corollary 12. *If $\|\xi\|_\infty < \infty$ then*

$$\Pr(A_n \textit{ has } o(n) \textit{ positive eigenvalues}) = \exp(-\Omega(n^2)). \tag{31}$$

Proof. At the cost of increasing the constant C, Theorem 11 also applies to functions f that are the difference of two 1-Lipschitz, convex functions (because we can apply Theorem 11 to each of the constituent functions and then apply a union bound). Consider the function

$$f(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x < 1 \\ 1 & x \geq 1. \end{cases} \tag{32}$$

Then $f(x) = \frac{1}{2}|x+1| - \frac{1}{2}|x|$ is the difference of two 1-Lipschitz, convex functions. Since f is bounded and $f(x) = 0$ for $x \leq 0$, it follows that if only $o(n)$ of the eigenvalues of A_n are positive then $X_n = o(1)$ (where $X_n = \frac{1}{n}\sum f(n^{-1/2}\lambda_i(A_n))$ as in Theorem 11. On the other hand, $\mathbb{E}X_n$ converges to a non-zero constant (call it c) because the empirical measure of $n^{-1/2}A_n$ converges weakly to the semicircle law [18]. Hence,

$$\begin{aligned} \Pr(A_n \text{ has } o(n) \text{ positive eigenvalues}) &\leq \Pr(|X_n - \mathbb{E}X_n| \geq c - o(1)) \\ &\leq \exp(-\Omega(n^2)) \end{aligned}$$

by Theorem 11. ∎

6. Upper bound

The main observation is that in the regime we are interested in (namely, eigenvalues or singular values of order $\omega(\sqrt{n})$, the probability of large eigenvalues can be controlled by a union bound over the potential eigenvectors.

Let $\mathcal{M}_k$ be the set of $n \times n$ matrices with rank at most k and Frobenius norm at most 1. Let $\mathcal{M}_k^+ \subset \mathcal{M}_k$ consist of those matrices that are symmetric and positive semidefinite.

Lemma 13. *For any symmetric matrix A,*

$$\left(\sum_{i=1}^{k} \max\{0, \lambda_i(A)\}^2\right)^{1/2} = \sup_{M \in \mathcal{M}_k^+} \langle A, M \rangle. \tag{33}$$

For any matrix A,

$$\left(\sum_{i=1}^{k} \sigma_i(A)^2\right)^{1/2} = \sup_{M \in \mathcal{M}_k} \langle A, M \rangle. \tag{34}$$

Proof. Let $UDU^T = A$ be an eigen-decomposition of A (where D is diagonal and U is orthogonal), and let $\tilde{D}$ be D but with all but the kth-largest diagonal entries set to zero. Define

$$M = \frac{U\tilde{D}U^T}{\|\tilde{D}\|_F} = \frac{U\tilde{D}U^T}{\left(\sum_{i=1}^{k} \lambda_i(A)^2\right)^{1/2}}. \tag{35}$$

Then $M \in \mathcal{M}_k^+$ and $\langle A, M \rangle = \|\tilde{D}\|_F = \left(\sum_{i=1}^{k} \lambda_i(A)^2\right)^{1/2}$. This proves one direction of the first claim.

For the other direction, take any $M \in \mathcal{M}_k^+$, and decompose A as $A_+ - A_-$, where A_+ and A_- are positive semi-definite and the non-zero eigenvalues of A_+ are the positive eigenvalues of A. Then

$$\langle A, M \rangle \le \langle A_+, M \rangle \le \|A_+\|_F \|M\|_F \le \sqrt{\sum_{i=1}^{k} \lambda_i(A_+)^2} = \sqrt{\sum_{i=1}^{k} \lambda_i(A)^2}. \tag{36}$$

This proves the first claim. The proof of the second claim is identical, but uses a singular value decomposition instead of an eigen-decomposition. □

Hence, in order to prove the upper bounds in Theorem 8, it suffices to control

$$\Pr\left(\sup_{M \in \mathcal{M}_k^+} \langle A, M \rangle > tn^\alpha\right). \tag{37}$$

The first step is to replace the supremum with a finite maximum.

6.1. **The net argument.**

Definition 14. *For a subset $\mathcal{N}$ of a metric space (X,d), we say that $\mathcal{N}$ is an ϵ-net of X if for every $x \in X$ there exists $y \in \mathcal{N}$ with $d(x,y) \le \epsilon$.*

Lemma 15. *Let $\mathcal{N} \subset \mathcal{M}_k$ be an ϵ-net (with respect to $\|\cdot\|_F$) for $\epsilon < \frac{1}{2}$. Then for any symmetric matrix A,*

$$\sup_{M \in \mathcal{M}_k} \langle A, M\rangle \le \frac{1}{1-2\epsilon} \sup_{N \in \mathcal{N}} \langle A, N\rangle. \tag{38}$$

Proof. Fix $M \in \mathcal{M}_k$, and choose $N \in \mathcal{N}$ with $\|N - M\|_F \le \epsilon$. Note that $N - M$ has rank at most $2k$, and it has at most k positive eigenvalues and k negative eigenvalues. Letting ϵM_0 and ϵM_1 be the positive and negative parts of $N - M$, we have $\|M_0\|_F \le \|N - M\|_F/\epsilon \le 1$ and (similarly $\|M_1\|_F \le 1$). In other words, we can decompose

$$M = N + \epsilon M_0 + \epsilon M_1 \tag{39}$$

with $N \in \mathcal{N}$ and $M_0, M_1 \in \mathcal{M}_k$. We continue this construction recursively: for every finite binary string v and matrix $M_v \in \mathcal{M}_k$, we can find $N_v \in \mathcal{N}$ and $M_{v0}, M_{v1} \in \mathcal{M}_k$ such that

$$M_v = N_v + \epsilon M_{v0} + \epsilon M_{v1}. \tag{40}$$

Recursing this construction m levels, it follows that (with S_m being the set of binary strings of length m and $|v|$ denoting the length of the string v)

$$M = \sum_{\ell=0}^{m-1} \sum_{v \in S_\ell} \epsilon^{-|v|} N_v + \epsilon^{-m} \sum_{v \in S_m} M_v. \tag{41}$$

Since $|S_m| = 2^m$ and each M_v has $\|M_v\|_F \le 1$, the remainder term converges to zero and we can continue this construction to the limit:

$$M = \sum_{\ell=0}^{\infty} \sum_{v \in S_\ell} \epsilon^{-|v|} N_v, \tag{42}$$

where the outer sum converges in Frobenius norm.

Taking the inner product with A, note that Cauchy-Schwarz and the convergence of the sum imply that the inner product and summation can be exchanged:

$$\langle M, A\rangle = \sum_{\ell=0}^{\infty} \sum_{v \in S_\ell} \epsilon^{-|v|} \langle N_v, A\rangle \le \sum_{\ell=0}^{\infty} |S_\ell| \epsilon^{-\ell} \sup_{N \in \mathcal{N}} \langle N, A\rangle = \frac{1}{1-2\epsilon} \sup_{N \in \mathcal{N}} \langle N, A\rangle. \tag{43}$$

□

The construction in Lemma 15 approximates the supremum over $M \in \mathcal{M}_k$, which is enough for most of what we will do. In some cases, we will want the supremum over $M \in \mathcal{M}_k^+$ instead, but that can be handled also:

Lemma 16. *Let $\mathcal{N} \subset \mathcal{M}_k$ and $\mathcal{N}^+ \subset \mathcal{M}_k^+$ be ϵ-nets (with respect to $\|\cdot\|_F$) for $\epsilon < \frac{1}{2}$ Then for any symmetric matrix A,*

$$\sup_{M \in \mathcal{M}_k^+} \langle A, M\rangle \le \sup_{N^+ \in \mathcal{N}^+} \langle A, N^+\rangle + \frac{2\epsilon}{1-2\epsilon} \sup_{N \in \mathcal{N}} \langle A, N\rangle. \tag{44}$$

Proof. Fix $M \in \mathcal{M}_k^+$ and choose $M_0 \in \mathcal{N}^+$ such that $\|M_0 - M\|_F \le \epsilon$. Then $\frac{M_0 - M}{\epsilon}$ has rank at most $2k$ and Frobenius norm at most 1. Hence, we can write $M_0 - M = \epsilon N_0 + \epsilon N_1$, where $N_0, N_1 \in \mathcal{M}_k$. It follows that

$$\langle A, M \rangle = \langle A, M_0 \rangle + \epsilon \langle A, N_0 \rangle + \epsilon \langle A, N_1 \rangle, \tag{45}$$

and we conclude by applying Lemma 15 to $\langle A, N_0 \rangle$ and $\langle A, N_1 \rangle$. □

We have shown that to approximate the supremum it suffices to take a good enough net. In order to put this together with a union bound, we need a bound on the size of a good net. Our starting point is the following basic bound in Euclidean space [19, Corollary 4.2.13]

Lemma 17. *The unit Euclidean ball in $\mathbb{R}^d$ admits an ϵ-net (with respect to the Euclidean metric) $\mathcal{N}$ satisfying $|\mathcal{N}| \le (3/\epsilon)^d$.*

Corollary 18. *There is a constant C such that for any $0 < \epsilon < 1$, there is an ϵ-net (with respect to Frobenius norm) for $\mathcal{M}_k$ of size at most $(C/\epsilon)^{2nk}$ and an ϵ-net for $\mathcal{M}_k^+$ of size at most $(C/\epsilon)^k$.*

Proof. Let $\tilde{\mathcal{N}}$ be an $(\epsilon/2)$-net for the set of $n \times k$ matrices with Frobenius norm at most one. Since this space is isometric to $\mathbb{R}^{nk}$ with the Euclidean norm, Lemma 17 implies that we can choose such a $\tilde{\mathcal{N}}$ with $|\tilde{\mathcal{N}}| \le (C/\epsilon)^{nk}$. Now let $\mathcal{N} = \{XY^T : X, Y \in \tilde{\mathcal{N}}\}$. Then $|\mathcal{N}| \le |\tilde{\mathcal{N}}|^2 \le (C/\epsilon)^{2nk}$.

It remains to show that $\mathcal{N}$ is an ϵ-net. Since $\|XY^T\|_F \le \|X\|_F \|Y\|_F$, it follows that every $N \in \mathcal{N}$ has $\|N\|_F \le 1$; also, each $N \in \mathcal{N}$ clearly has rank at most k. Now choose an arbitrary $M \in \mathcal{M}_k$ and write $M = AB^T$ for $n \times k$ matrices A and B of Frobenius norm at most 1 (for example, this can be done using a singular value decomposition). Choose $X, Y \in \tilde{\mathcal{N}}$ with $\|X - A\|_F \le \frac{\epsilon}{2}$ and $\|Y - B\|_F \le \frac{\epsilon}{2}$. Then

$$\begin{aligned}
\|XY^T - M\|_F &\le \|XY^T - AY^T\|_F + \|AY^T - AB^T\|_F \\
&\le \|X - A\|_F + \|Y^T - B^T\|_F \\
&\le \epsilon.
\end{aligned}$$

To construct an ϵ-net of $\mathcal{M}_k^+$, take $\tilde{\mathcal{N}}$ be as above and let $\mathcal{N} = \{XX^T : X \in \tilde{\mathcal{N}}\}$. Then $|\mathcal{N}| \le |\tilde{\mathcal{N}}|$, and the proof that $\mathcal{N}$ is an ϵ-net of $\mathcal{M}_k^+$ is essentially the same as the proof above, the only change being that every $M \in \mathcal{M}_k^+$ can be written as $M = AA^T$ for an $n \times k$ matrix A of Frobenius norm at most 1. □

Applying a union bound over these nets gives the main result of this section: singular values and eigenvalues of A can be controlled in terms of the deviations of linear functions of A. The main point here is that (as we will show in the next section) if $t \gg \sqrt{n}$ then the $O(nk \ln \frac{1}{\epsilon})$ terms are negligible compared to the other terms.

Proposition 19. *Let A be a symmetric $n \times n$ random matrix with i.i.d. entries. For any integer $k \ge 1$, any $0 < \epsilon < \frac{1}{2}$, and any $t > 0$,*

$$\ln \Pr\left(\sum_{i=1}^k \sigma_i^2(A) > t\right) \le \sup_{M \in \mathcal{M}_k} \ln \Pr\left(\langle A, M \rangle \ge (1 - 2\epsilon)t\right) + O(nk \ln \frac{1}{\epsilon}). \tag{46}$$

If, in addition, $\|\xi\|_\infty < \infty$ *then*

$$\ln \Pr\left(\sum_{i=1}^k \lambda_i^2(A) > t\right) \le \sup_{M^+ \in \mathcal{M}_k^+} \ln \Pr\left(\langle A, M^+\rangle \ge (1-\sqrt{\epsilon})t\right) + \sup_{M \in \mathcal{M}_k} \ln \Pr\left(\langle A, M\rangle \ge \frac{1-2\epsilon}{2\sqrt{\epsilon}} t\right) + O(nk \ln \frac{1}{\epsilon}) + O(n^2)$$

Proof. For the first inequality, let $\mathcal{N}$ be an ϵ-net for $\mathcal{M}_k$ according to Corollary 18. By Lemma 13 and Lemma 15

$$\Pr\left(\sum_{i=1}^k \sigma_i^2(A) > t\right) = \Pr\left(\sup_{M \in \mathcal{M}_k} \langle A, M\rangle > t\right) \le \Pr\left(\max_{N \in \mathcal{N}} \langle A, N\rangle > (1-2\epsilon)t\right).$$

By a union bound,

$$\Pr\left(\max_{N \in \mathcal{N}} \langle A, N\rangle > (1-2\epsilon)t\right) \le \sum_{N \in \mathcal{N}} \Pr\left(\langle A, N\rangle > (1-2\epsilon)t\right) \le |\mathcal{N}| \sup_{M \in \mathcal{M}_k} \Pr\left(\langle A, M\rangle > (1-2\epsilon)t\right),$$

which, by our bound on $|\mathcal{N}|$, completes the proof of the first claim.

The second claim is similar: by Lemma 16, $\sup_{M^+ \in \mathcal{M}_k^+} \langle A, M^+\rangle > t$ implies that either $\max_{N^+ \in \mathcal{N}^+} \langle A, N^+\rangle > (1-\sqrt{\epsilon})t$ or $\max_{N \in \mathcal{N}} \langle A, N\rangle > \frac{1-2\epsilon}{2\sqrt{\epsilon}} t$. Hence,

$$\Pr\left(\sup_{M^+ \in \mathcal{M}_k} \langle A, M^+\rangle > t\right) \le \Pr\left(\max_{N^+ \in \mathcal{N}^+} \langle A, N^+\rangle > (1-\sqrt{\epsilon})t\right) + \Pr\left(\max_{N \in \mathcal{N}} \langle A, N\rangle > \frac{1-2\epsilon}{2\sqrt{\epsilon}} t\right),$$

and applying the union bound as before yields

$$\ln \Pr\left(\sup_{M^+ \in \mathcal{M}_k} \langle A, M^+\rangle > t\right) \le \sup_{M^+ \in \mathcal{M}_k^+} \ln \Pr\left(\langle A, M^+\rangle \ge (1-\sqrt{\epsilon})t\right) + \sup_{M \in \mathcal{M}_k} \ln \Pr\left(\langle A, M\rangle \ge \frac{1-2\epsilon}{2\sqrt{\epsilon}} t\right) + O(nk \ln \frac{1}{\epsilon}).$$

To complete the proof, we apply Lemma 13 along with the fact, by Corollary 12, that with probability at least $\exp(-\Omega(n^2))$, A has at least k positive eigenvalues. ☐

Note that we will not actually use the second part of Proposition 19. We mention it only to point out that since

$$\sup_{M^+ \in \mathcal{M}_k^+} \ln \Pr\left(\langle A, M^+\rangle \ge t\right) \le \sup_{M \in \mathcal{M}_k} \ln \Pr\left(\langle A, M\rangle \ge t\right), \tag{47}$$

we can in principle get a better bound on the eigenvalues than for the singular values. The issue is that we do not know how to exploit the additional information that we are testing A against a positive semidefinite matrix.

6.2. **Hoeffding-type argument.** Using a Hoeffding-type argument, we can get a sharp upper bound on

$$\sup_{M \in \mathcal{M}_k} \ln \Pr\left(\langle A, M\rangle \geq t\right) \tag{48}$$

for any k and any t (in fact, the sharp upper bound turns out not to depend on k).

Lemma 20. *If ξ is subgaussian then*

$$4 \sup_{s \in \mathbb{R}} \frac{\Lambda_\xi(s)}{s^2} = \left(\inf_{s \in \mathbb{R}} \frac{\Lambda_\xi^*(u)}{u^2} \right)^{-1} < \infty. \tag{49}$$

Proof. The fact that $\sup_{s\in\mathbb{R}} \frac{\Lambda(s)}{s^2} < \infty$ is the definition of subgaussianity. To show the claimed identity, let $L = \sup_{t \in \mathbb{R}} \frac{\Lambda_\xi(t)}{t^2}$ and define $M_L(s) = Ls^2$. Clearly, $\Lambda(s) \leq M_L(s)$ for all $s \in \mathbb{R}$. It follows that $\Lambda^*(u) \geq M_L^*(u) = \frac{u^2}{4L}$; in other words,

$$\frac{\Lambda^*(u)}{u^2} \geq \frac{M_L^*(u)}{u^2} = \frac{1}{4L} \tag{50}$$

for all u. This shows that

$$4 \sup_{s \in \mathbb{R}} \frac{\Lambda(s)}{s^2} \geq \left(\inf_{u \in \mathbb{R}} \frac{\Lambda^*(u)}{u^2} \right)^{-1}. \tag{51}$$

For the other direction, suppose that for some L' we have $\Lambda^*(u) \geq \frac{u^2}{4L'} = M^{1/(4L)'}(u)$ for every u. Then (since Λ is convex) $\Lambda(t) = \Lambda^{**}(t) \leq M^*_{1/(4L')}(t) = L't^2$ for every t. The definition of L ensures that $L' \geq L$, and this shows the other direction of the claim. ☐

Proposition 21. *Let ξ be a random variable with globally finite moment-generating function, and define*

$$\Lambda(s) = \ln \mathbb{E} \exp(s\xi) \tag{52}$$

to be the cumulant-generating function of ξ. Let A be a symmetric random matrix with zero diagonal, and with upper-diagonal elements distributed independently according to ξ. Define $\ell^* = \sup_{s>0} \frac{\Lambda(s)}{s^2}$. Then

$$\sup_{\|M\|_F \leq 1} \Pr(\langle A, M\rangle > t) \leq \exp\left(-\frac{t^2}{8 \sup_{s>0} \frac{\Lambda(s)}{s^2}} \right) = \exp\left(-\frac{t^2}{2} \inf_{s>0} \frac{\Lambda^*(s)}{s^2} \right). \tag{53}$$

Proof. Since $\langle A, M\rangle = \langle A, (M + M^T)/2\rangle$ and since $\|(M + M^T)/2\|_F \leq \|M\|_F$, it suffices to consider only symmetric matrices M. Let $m = \frac{n}{n-1}2$ and let $\xi_1, \dots, \xi_m$ be the upper-diagonal elements of A, in any order. Let $\|M\| \leq 1$ be symmetric, with upper-diagonal entries $a_1, \dots, a_m$. Then $\langle A, M\rangle = 2\sum_{i=1}^m a_i \xi_i$, and so (for any $s > 0$)

$$\begin{aligned}\Pr(\langle A, M\rangle > t) &= \Pr\left(\sum a_i\xi_i > t/2\right)\\ &= \Pr\left(e^{s\sum a_i\xi_i} > e^{st/2}\right)\\ &\le e^{-st/2}\mathbb{E}e^{s\sum a_i\xi_i}\\ &= \exp\left(\sum_i \Lambda(sa_i) - st/2\right),\end{aligned}$$

where the inequality follows from Markov's inequality. Now, $\sum_{i=1}^m a_i^2 \le \frac{1}{2}\|M\|_F^2 \le \frac{1}{2}$, and so if we set $\ell^* = \sup_{r>0}\frac{\Lambda(r)}{r^2}$ then

$$\sum_i \Lambda(sa_i) = \sum_i \frac{\Lambda(sa_i)}{(sa_i)^2}(sa_i)^2 \le s^2\sum_i \ell^* a_i^2 \le \frac{s^2\ell^*}{2}. \tag{54}$$

Hence,

$$\Pr(\langle A, M\rangle > t) \le \exp\left(\frac{s^2\ell^*}{2} - \frac{st}{2}\right), \tag{55}$$

and the first claim follows by optimizing over s.

The second claim follows immediately from Lemma 20. □

Putting Proposition 21 together with Proposition 19, we arrive at the following upper bound for singular values:

Corollary 22. *Let A be a symmetric $n\times n$ random matrix with i.i.d. upper diagonal entries. Assuming that the entries are subgaussian and have cumulant-generating function Λ, let $L = \inf_{s\in\mathbb{R}}\frac{\Lambda^*(s)}{s^2}$. Then for any integer k and any $t > 0$, if $t^2L > 2nk$ then*

$$\ln\Pr\left(\sqrt{\sum_{i=1}^k s_i^2(A)} > t\right) \le -\frac{t^2L}{2} + O\left(nk\ln\frac{t^2L}{nk}\right). \tag{56}$$

Proof. We combine Proposition 21 and Proposition 19, setting $\epsilon = \frac{nk}{t^2L}$ (which is less than $\frac{1}{2}$ by our assumption on C). This yields an upper bound of

$$-\frac{t^2L}{2} + O\left(nk + nk\ln\frac{t^2L}{nk}\right), \tag{57}$$

and the nk term can be absorbed in the final term. □

6.3. **Lower bound.** In this section, we give a lower bound that matches the upper bound of Corollary 22 whenever $\sqrt{n} \ll t \ll n$. The starting point is the lower bound of Cramér's theorem [20, Theorem 27.3]

Theorem 23. *Let ξ be a mean-zero random variable with everywhere-finite cumulant-generating function Λ_ξ. Let $\xi_1, \dots, \xi_m$ be independent copies of ξ. Then*

$$\frac{1}{m}\ln\Pr\left(\sum_{i=1}^m \xi_i > mt\right) \to -\Lambda^*(t) \tag{58}$$

as $m \to \infty$.

Proposition 24. *In the setting of Corollary 22, suppose in addition that the function $s \mapsto s^{-2}\Lambda^*(s)$ achieves its minimum at some finite $s \in \mathbb{R}$. Then for any $1 \ll t \ll n^2$ and for any $w_1, \dots, w_k > 0$, we have*

$$\ln \Pr\left(\sum_{i=1}^k w_i \sigma_i(A_n) > |w|\sqrt{t}\right) \geq -\frac{tL}{2} - o(t). \tag{59}$$

If $s \mapsto s^{-2}\Lambda^*(s)$ achieves its minimum at some $s \geq 0$, then for any $1 \ll t \ll n^2$ and for any $w_1, \dots, w_k > 0$, we have

$$\ln \Pr\left(\sum_{i=1}^k w_i \lambda_i(A_n) > |w|\sqrt{t}\right) \geq -\frac{tL}{2} - o(t). \tag{60}$$

Choosing an arbitrary $w_1, \dots, w_k$ and applying the Cauchy-Schwarz inequality, Proposition 24 implies the same lower bounds on $\ln \Pr(\sum_i \sigma_i^2(A_n) > \sqrt{t})$ and $\ln \Pr(\sum_i \lambda_i^2(A_n) > \sqrt{t})$. In particular, it really is a lower bound that matches the upper bound of Corollary 22.

Proof. Fix t and assume that $\frac{\Lambda^*(s)}{s^2}$ achieves its minimum at $s_* \in \mathbb{R}$. Actually, we will assume $s_* \neq 0$; the case $s_* = 0$ is easily handled by replacing s_* with $\epsilon > 0$ everywhere, and then sending $\epsilon \to 0$. Fix $w_1, \dots, w_k$ and assume $\sum_i w_i^2 = t$; because the statement of the proposition is homogeneous in w, this is without loss of generality. Now choose the smallest integers $\ell_1, \dots, \ell_k$ so that $\ell_i - 1 \geq \frac{w_i}{s_*}$. We write $|\ell|^2$ for $\sum_i \ell_i^2$, and note that $|\ell|^2 \geq \frac{1}{s_*^2}\sum_i w_i^2 = \frac{t}{s_*^2}$, meaning that $1 \ll |\ell|^2 \ll n^2$.

Let M be a block-diagonal matrix, whose non-zero entries are all equal to s_*, appearing in blocks of size $\ell_i \times \ell_i$ for $i = 1, \dots, k$. (The fact that $\sum_i \ell_i \leq \sqrt{k}|\ell| \ll n$ implies that these blocks do indeed fit into an $n \times n$ matrix.) Then M has rank k, and the singular values of M are $s_*\ell_i$ for $i = 1, \dots, k$; note that our choices of ℓ_i ensure that $w_i \leq \sigma_i(M) \leq w_i + 2s_*$. Moreover, if we set $m = \sum_i \frac{\ell_i(\ell_i - 1)}{2}$ (which is also an integer, and counts the number of non-zero upper-diagonal elements of M) then $\langle A, M\rangle$ is equal in distribution to $2s_* \sum_{i=1}^m \xi_i$. Hence,

$$\Pr\left(\langle A, M\rangle > t\right) = \Pr\left(\operatorname{sgn}(s_*)\sum_{i=1}^m \xi_i > \frac{t}{2|s_*|}\right). \tag{61}$$

Now, $m = \frac{1}{2}|\ell|^2 - \frac{1}{2}\sum_i \ell_i$, while on the other hand

$$\frac{t}{s_*^2} = \frac{\sum_i w_i^2}{s_*^2} \leq \sum_i (\ell_i - 1)^2 = |\ell|^2 - 2\sum_i \ell_i + 2k. \tag{62}$$

Since $\sum_i \ell_i \geq |\ell| \gg 1$, we have $\frac{t}{2s_*^2} \leq m$ for sufficiently large n. Going back to our probability estimates, we have

$$\begin{aligned}
\ln \Pr\left(\langle A, M\rangle > t\right) &= \ln \Pr\left(\operatorname{sgn}(s_*) \sum_{i=1}^m \xi_i > \frac{t}{2|s_*|}\right) \\
&\geq \ln \Pr\left(\operatorname{sgn}(s_*) \sum_{i=1}^m \xi_i > m|s_*|\right) \\
&= -m\Lambda^*(s_*) + o(m) \\
&= -\frac{t\Lambda^*(s_*)}{2s_*^2} - o(t),
\end{aligned}$$

where the second-last equality follows by Cramér's theorem (applied to the random variables $-\xi_i$ in case $s_* < 0$). By the Cauchy-Schwarz inequality we have

$$\begin{aligned}
\langle A, M\rangle \leq \sum_{i=1}^k \sigma_i(A)\sigma_i(M) \leq \sum_{i=1}^k \sigma_i(A)(w_i + 2s_*) \\
\leq \sum_{i=1}^k \sigma_i(A) w_i + 2s_*\sqrt{k}\sqrt{\sum_{i=1}^k \sigma_i^2(A)},
\end{aligned} \tag{63}$$

and hence

$$\Pr\left(\langle A, M\rangle > t\right) \leq \Pr\left(\sum_{i=1}^k \sigma_i(A) w_i > t - t^{2/3}\right) + \Pr\left(\sum_{i=1}^k \sigma_i^2(A) > \frac{t^{4/3}}{4s_*^2 k}\right). \tag{64}$$

By Corollary 22, the second probability is of order $\exp(-\Omega(t^{4/3}))$, and hence

$$\ln \Pr\left(\sum_{i=1}^k \sigma_i(A) w_i > t - t^{2/3}\right) \geq (1 - o(1)) \ln \Pr\left(\langle A, M\rangle > t\right) \geq -\frac{t\Lambda^*(s_*)}{2s_*^2} - o(t). \tag{65}$$

Substituting in t in place of $t - t^{2/3}$, the extra error term can be absorbed in the $o(t)$ term. This proves the first claim.

For the second claim, simply note that if $s_* > 0$ then the matrix M is positive semi-definite. Denoting $\lambda_i^+(A) = \max\{0, \lambda_i(A)\}$, we replace (63) by

$$\begin{aligned}
\langle A, M\rangle \leq \sum_{i=1}^k \lambda_i^+(A)\lambda_i(M) \leq \sum_{i=1}^k \lambda_i^+(A)(w_i + 2s_*) \\
\leq \sum_{i=1}^k \lambda_i^+(A) w_i + 2s_*\sqrt{k}\sqrt{\sum_{i=1}^k \sigma_i^2(A)},
\end{aligned}$$

and the rest of the proof proceeds as before. □

There are a few extra useful facts that we can extract from the proof of Proposition 24, namely that we have explicit candidates for extremal eigenvectors and singular vectors. We will state these just for the smallest eigenvector, but of course they also hold in other situations.

Corollary 25. *Assume that $s \mapsto s^{-2}\Lambda^*(s)$ achieves its minimum at some $s_* \geq 0$. For $1 \ll t \ll n^2$, let $\ell = \lfloor \sqrt{t}/s_* \rfloor$ and define $v \in \mathbb{R}^n$ by $v_1, \dots, v_\ell = s_*^{1/2} t^{-1/4}$ and $v_{\ell+1}, \cdots, v_n = 0$. Then $|v| \leq 1$ and*

$$\ln \Pr(v^T A_n v \geq \sqrt{t}) \geq -\frac{tL}{2} - o(t). \tag{66}$$

Moreover, if $\tilde{v} = v - \frac{\sum_i v_i}{n}\mathbf{1}$ then $|\tilde{v}| \leq 1 + o(1)$ and

$$\ln \Pr(\tilde{v}^T A_n \tilde{v} \geq \sqrt{t}) \geq -\frac{tL}{2} - o(t). \tag{67}$$

Proof. The first claim is immediate from the proof of Proposition 24, because the M that we constructed in that proof is exactly vv^T.

For the second claim, note that $|v - \tilde{v}| = O(t^{1/4} n^{-1/2}) = o(1)$. It follows that $\|vv^T - \tilde{v}\tilde{v}^T\|_F = o(1)$, and then Hoeffding's inequality implies that

$$\ln \Pr(|\tilde{v}^T A_n \tilde{v} - v^T A_n v| \geq \sqrt{t}) \leq -\omega(t). \tag{68}$$

□

6.4. **The LDP.** Putting together Corollary 22 and Proposition 24, we complete the proof of the LDP (Theorem 8). Take a sequence m_n satisfying $\sqrt{n} \ll m_n \ll n$, and set $X = \frac{1}{m_n}(\sigma_1(A_n), \dots \sigma_k(A_n))$. Let $E \subset \mathbb{R}^k$ be any closed set, and let $t = \inf_{x \in E} |x|$. Then $\frac{1}{m_n}(\sigma_1(A_n), \dots, \sigma_k(A_n)) \in E$ implies that $\sum \sigma_i^2(A_n) > m_n^2 t^2$. By Corollary 22,

$$\begin{aligned}\ln \Pr(X \in E) &\leq \ln \Pr\left(\sum_{i=1}^k \sigma_i^2(A_n) > m_n^2 t^2\right) \\ &\leq -\frac{m_n^2 t^2 L}{2} + O\left(n \ln \frac{m_n^2}{n}\right) = -\frac{m_n^2 t^2 L}{2} + o(m_n^2).\end{aligned}$$

On the other hand, if $E \subset \mathbb{R}^k$ is open, then choose any $w \in E$. Since E is open, there is some $\epsilon > 0$ so that if $\langle x, w \rangle \geq |w|^2$ and $|x|^2 \leq |w|^2 + \epsilon$ then $x \in E$. Now, Proposition 24 implies that

$$\ln \Pr\left(\langle X, w \rangle \geq |w|^2\right) = \ln \Pr\left(\sum_i \sigma_i(A_n) w_i \geq m_n |w|^2\right) \geq -\frac{m_n^2 |w|^2 L}{2} - o(m_n^2) \tag{69}$$

On the other hand, Corollary 22 implies that

$$\begin{aligned}\ln \Pr\left(|X|^2 > |w|^2 + \epsilon\right) &= \ln \Pr\left(\sum_i \sigma_i^2(A_n) \geq m_n^2(|w|^2 + \epsilon)\right) \\ &\leq -\frac{m_n^2(|w|^2 + \epsilon) L}{2} - o(m_n^2).\end{aligned}$$

In particular, $\Pr(|X|^2 > |w|^2 + \epsilon)$ is dominated by $\Pr(\langle X, w \rangle \geq |w|^2)$, implying that

$$\ln \Pr(X \in E) \geq \ln \Pr\left(\langle X, w \rangle \geq |w|^2 \text{ and } |X|^2 \leq |w|^2 + \epsilon\right) \geq -\frac{m_n^2 |w|^2 L}{2} - o(m_n^2). \tag{70}$$

Since this holds for arbitrary $w \in E$, it implies the lower bound in the LDP.

The second part of Theorem 8 follows the exact same argument, only it uses the second part of Proposition 24.

6.5. **The case of $\mathcal{G}(m,n)$.** We next consider the case of Theorem 10. The first observation is that if $p \le \frac{1}{2}$ if and only if $\Lambda^*(s)/s^2$ achieves its minimum at some non-negative s.

Lemma 26. *If $\xi = -p$ with probability $1-p$ and $\xi = 1-p$ with probability p then Λ^* (the convex conjugate of ξ's cumulant generating function) satisfies*

$$\inf_{s\in\mathbb{R}} \frac{\Lambda^*(s)}{s^2} = \frac{\ln\frac{1-p}{p}}{1-2p}, \tag{71}$$

and the minimum is uniquely attained at $s = 1-2p$.

Proof. We recall that $\Lambda^*(s) = D(p+s,p)$ where

$$D(q,p) = q\ln\frac{q}{p} + (1-q)\ln\frac{1-q}{1-p} \tag{72}$$

(with the convention that $D(q,p) = +\infty$ for $q \notin (0,1)$). Note that $D(q,p)$ is non-negative, convex, and has a double-root at $q = p$. Fix p and define

$$L(q) = \frac{D(q,p)}{(q-p)^2} = \frac{\Lambda^*(q-p)}{(q-p)^2} \tag{73}$$

(defined by continuity at $q = p$); our task is then to minimize L. We compute

$$L'(q) = -\frac{(p+q)\ln\frac{q}{p} + (2-p-q)\ln\frac{1-q}{1-p}}{(q-p)^3} =: -\frac{F(q)}{(q-p)^3}. \tag{74}$$

Then

$$\begin{aligned} F'(q) &= \ln\frac{q}{p} - \ln\frac{1-q}{1-p} + \frac{p}{q} - \frac{1-p}{1-q} \\ F''(q) &= (q-p)\left(\frac{1}{q^2} - \frac{1}{(1-q)^2}\right). \end{aligned}$$

In particular, F'' has exactly two roots on $(0,1)$: at $q = \frac{1}{2}$ and at $q = p$ (counting with multiplicity in case $p = \frac{1}{2}$). It follows that F has at most 4 roots on $(0,1)$. On the other hand, we can easily see that $F(p) = F'(p) = F''(p) = F(1-p) = 0$. Hence, $F(q)$ has a triple-root at $q = p$ and a single root at $q = 1-p$, and no other roots. Since $q = p$ is only a triple-root, $L'(p) \ne 0$, and it follows that $q = 1-p$ is the only root of $L'(q)$. It follows that $L(q)$ is minimized at either $q = 0$, $q = 1$, or $q = 1-p$. The possible minimum values are therefore

$$x := p^{-2}\ln\frac{1}{1-p}, \qquad y := (1-p)^{-2}\ln\frac{1}{p}, \qquad \text{or } z := \frac{\ln\frac{1-p}{p}}{1-2p}. \tag{75}$$

We will show that z is the smallest one. By symmetry in p and $1-p$, it suffices to show that $z \le x$ for all p. Now,

$$p^2(1-2p)(z-x) = p^2\ln\frac{1-p}{p} + (1-2p)\ln(1-p) = (1-p)^2\ln(1-p) - p^2\ln p. \tag{76}$$

Let $f(p) = (1-p)^2 \ln(1-p) - p^2 \ln p$, and we need to show that $f(p) < 0$ for $0 < p < \frac{1}{2}$ and $f(p) > 0$ for $\frac{1}{2} < p < 1$. In fact, since $f(p) = -f(1-p)$, it suffices to show only one of these. Finally, note that $f(0) = f(\frac{1}{2}) = 0$, and $f''(p) > 0$ for $0 < p < \frac{1}{2}$, and it follows that $f(p) < 0$ for $0 < p < \frac{1}{2}$. □

To complete the proof of Theorem 10, it is enough to show that the upper bound of Corollary 22 and the lower bound of Proposition 24 still hold in this setting; then the proof of the LDP proceeds exactly as in the proof of Theorem 8. Checking Corollary 22 is trivial: recalling that A_n is the centered adjacency matrix of $\mathcal{G}(n,m)$ for $|m - p\binom{n}{2}| = O(1)$, we let $\tilde{A}_n$ be the centered adjacency matrix of $\mathcal{G}(n,p)$. Note that the distribution of A_n is equal to the distribution of $\tilde{A}_n$, conditioned on the event that $\tilde{A}_n$ has exactly m positive entries on the upper diagonal; call this event E. By Stirling's approximation, $\Pr(E) = \Omega(n^{-1})$, and it follows that for any event F,

$$\Pr(A_n \in F) = \Pr(\tilde{A}_n \in F \mid E) \le \frac{\Pr(A_n \in F)}{\Pr(E)} \le O(n \Pr(\tilde{A}_n \in F)). \tag{77}$$

In other words, $\ln \Pr(A_n \in F) \le \ln \Pr(\tilde{A}_n \in F) + O(\ln n)$, and so Corollary 22 immediately implies the same upper bound for $\mathcal{G}(n,m)$.

For the lower bound, we need to look into the proof of Proposition 24. Recall that in the proof of Proposition 24, we constructed a matrix M with $O(t) = o(n^2)$ non-zero entries, all of which had the same value. For the $\mathcal{G}(n,p)$ adjacency matrix $\tilde{A}_n$, $\langle \tilde{A}_n, M\rangle$ has a (scaled and translated) binomial distribution; for the $\mathcal{G}(n,m)$ adjacency matrix A_n, $\langle A_n, M\rangle$ has a (scaled and translated) hypergeometric distribution. Now, if $H_{k,n,q}$ denotes a hypergeometric random variable with population size n, k successes, and q trials; and if $B_{p,q}$ denotes a binomial random variable with success probability p and q trials; then one easily shows using Stirling's approximation that

$$|\ln \Pr(H_{k,n,q} = r) - \ln \Pr(B_{k/n,q} = r)| = O(q^2/n). \tag{78}$$

In the setting of Proposition 24, the number of trials q is the number of non-zero elements in M, and since $q^2/n = O(t^2/n) = o(t)$, we have

$$\ln \Pr(\langle A_n, M\rangle > t) \ge \ln \Pr(\langle \tilde{A}_n, M\rangle > t) - o(t). \tag{79}$$

With this lower bound, we can follow the rest of the proof of Proposition 24 to complete the proof of Theorem 10.

7. Proof of Theorem 9

Next, we consider the case that $\frac{\Lambda^*(s)}{s^2}$ does not achieve its infimum at any $s > 0$, and we construct an example showing that taking $s \to 0$ does not yield the sharp bound. The basic idea is to use the first part of Lemma 13, by producing a positive semi-definite matrix M and giving a lower bound on the tails of $\langle A, M\rangle$. The main challenge is to find a good matrix satisfying the positive definiteness constraint: in Proposition 24 we chose a matrix taking only one non-zero value, specifically, $s_* \in \operatorname{argmin} \frac{\Lambda^*(s)}{s^2}$. The issue, of course, is that if s_* is negative then such matrix cannot be positive semi-definite. Instead, we will construct a rank-1 matrix taking four different non-zero values.

Consider a sequence $a_1, \dots, a_n$ whose non-zero elements take m different values, $\alpha b_1, \dots, \alpha b_m$, with αb_i repeated $\tilde{m}_i = \beta m_i(1 + o(1))$ times respectively (the addition of the error term just allows us to deal with the fact that matrices have integer numbers of rows and columns). We will think of m_i and b_i as being fixed, while α and β depend on the tail bound that we want to show, with α being small and β being large. Then for any $t = \sum_{i=1}^m t_i$,

$$\Pr\left(\sum_i a_i \xi_i > t\right) \geq \prod_{i=1}^m \Pr\left(\sum_{j=1}^{\lceil \tilde{m}_i \rceil} \xi_j > t/(\alpha b_i)\right) \tag{80}$$

and so Theorem 23 implies that if $\frac{t_i}{\alpha \beta m_i b_i} = \Theta(1)$ then

$$\ln \Pr\left(\sum_i a_i \xi_i > t\right) \geq -\beta \sum_i m_i \Lambda^*\left(\frac{t_i}{\alpha \beta m_i b_i}\right) - o\left(\beta \sum_i m_i\right) \tag{81}$$

Our goal will be to choose the parameters m_i, b_i, α, β, and t_i to make the right hand side large. First, we will treat m_i and b_i as given, and optimize over t_i, α, and β. We will enforce the constraints $\sum_i t_i = t$ and $\sum_i a_i^2 = \alpha^2 \beta \sum_i m_i b_i^2 = 2$.

Define

$$\begin{aligned}
\beta &= t^2 \frac{\sum_i m_i b_i^2}{2\left(\sum_i m_i b_i \Lambda'(b_i)\right)^2} \\
\alpha &= 2\left(\beta \sum_i m_i b_i^2\right)^{-1/2} = \frac{\sum_i m_i b_i \Lambda'(b_i)}{t \sum_i m_i b_i^2} \\
t_i &= \alpha \beta m_i b_i \Lambda'(b_i).
\end{aligned}$$

With these choices, we have

$$\alpha^2 \beta = \frac{2}{\sum_i m_i b_i^2}, \tag{82}$$

meaning that

$$\sum_i a_i^2 = \alpha^2 \beta \sum_i m_i b_i^2 = 2 \tag{83}$$

and

$$\sum_i t_i = \alpha \beta \sum_i m_i b_i \Lambda'(b_i) = t. \tag{84}$$

(These turn out to be the optimal choices of α, β, and t, although we do not need to show this, since any choice will give us a bound.) Plugging these parameters into (81), we obtain

$$\ln \Pr\left(\sum_i a_i \xi_i > t\right) \geq -\frac{t^2}{2} \cdot \frac{\sum_i m_i b_i^2 \cdot \sum_i m_i \Lambda^*(\Lambda'(b_i))}{\left(\sum_i m_i b_i \Lambda'(b_i)\right)^2} - o(t^2), \tag{85}$$

where the $o(t^2)$ term depends on the parameters m_i and b_i.

Next, we will define the parameters m_i and b_i. Take $\epsilon, \delta > 0$, and define

$$
\begin{aligned}
m_1 &= \frac{1}{\epsilon^2} & b_1 &= \epsilon \\
m_2 &= 2\frac{\epsilon}{\delta^3} & b_2 &= -\delta \\
m_3 &= \frac{\epsilon^4}{\delta^6} & b_3 &= \frac{\delta^2}{\epsilon},
\end{aligned}
$$

and note that it is possible to define a positive semi-definite integral kernel taking the value $b_i/2$ on a set of measure $2m_i$, simply by starting with a function taking the values $\sqrt{\epsilon}$ and $-\delta/\sqrt{\epsilon}$ on sets of size $1/\epsilon$ and ϵ/δ^3 respectively, and then taking the outer product of that function with itself. It follows that if ϵ and δ are fixed and β is large (and α is arbitrary), then we can define a rank-1 p.s.d. matrix (M, say) with $(1 + o(1))2\beta m_i$ entries taking the value $\alpha b_i/2$; note that $\|M\|_F^2 = \frac{1+o(1)}{2}\alpha\beta^2 \sum_i m_i = 1 + o(1)$. Since A is a symmetric matrix with ξ on the upper diagonal, this will yield

$$\langle A, M\rangle = \sum_i a_i \xi_i \tag{86}$$

where (a_i) is a sequence containing $(1 + o(1))\beta m_i$ copies of αb_i.

We will first choose a small δ and then choose a smaller ϵ. The error terms in the following analysis are taking this into account, so for example we may write $\epsilon^2\delta^{-k} = o(\epsilon)$ no matter how large k is. Our next task is to compute the various expressions in (85), in terms of ϵ and δ. Before doing so, we observe some basic properties of the Legendre transform.

Lemma 27. *Assume that f is convex and differentiable and* $\lim_{x\to\infty} \frac{f(x)}{x^2} = 0$. *Then* $\lim_{x\to\infty} \frac{f^*(f'(x))}{x^2} = 0$.

Proof. Fix x and let $y = f'(x)$. By the definition of f^*, we can write

$$f^*(y) = \sup_z \{zy - f(z)\}, \tag{87}$$

and note that the supremum is attained at $x = z$ (because the derivative is zero, and the expression being supremized is concave). Hence,

$$f^*(f'(x)) = xf'(x) - f(x). \tag{88}$$

Convexity of f implies that f' is non-decreasing, and so $f(x) = o(x^2)$ implies that $f'(x) = o(x)$ as $x \to \infty$. Hence, $f^*(f'(x)) = xf'(x) - f(x) = o(x^2)$. □

Lemma 28. *If f is convex with $f(0) = f'(0) = 0$ and $f''(0) > 0$, and if both f and f^* are $\mathcal{C}^4$ in a neighborhood of 0, then*

$$f^*(f'(\epsilon)) = f''(0)\frac{\epsilon^2}{2} + ((f^*)'''(0)(f'')^3(0) + 3f'''(0))\frac{\epsilon^3}{6} + O(\epsilon^4) \tag{89}$$

as $\epsilon \to 0$

Proof. This is nothing but Taylor's theorem and a computation. Setting $g = f^*$, we compute

$$\frac{d}{d\epsilon} g(f'(\epsilon)) = g'(f'(\epsilon))f''(\epsilon), \tag{90}$$

and then

$$\frac{d^2}{d\epsilon^2} g(f'(\epsilon)) = g''(f'(\epsilon))(f''(\epsilon))^2 + g'(f'(\epsilon))f'''(\epsilon), \tag{91}$$

and finally

$$\frac{d^3}{d\epsilon^3} g(f'(\epsilon)) = g'''(f'(\epsilon))(f''(\epsilon))^3 + 3g''(f'(\epsilon))f''(\epsilon)f'''(\epsilon) + g'(f'(\epsilon))f''''(\epsilon). \tag{92}$$

Our assumptions on f ensure that $g'(0) = 0$, and hence the first-order term vanishes, the second-order term at $\epsilon = 0$ becomes

$$g''(0)(f''(0))^2, \tag{93}$$

and the third-order term at $\epsilon = 0$ becomes

$$g'''(0)(f''(0))^3 + 3g''(0)f''(0)f'''(0). \tag{94}$$

Finally, note that $g''(0)f''(0) = 1$. □

Note that Λ satisfies the assumptions on f in Lemmas 27 (because we assumed that $\Lambda(s) = o(s^2)$) and 28 (because every cumulant-generating function defined on a neighborhood of zero is $\mathcal{C}^\infty$ in a neighborhood of zero). Note that Λ and Λ^* both have a second-order root at zero. Define

$$L = \Lambda''(0) > 0. \tag{95}$$

Expanding out the parameters in (85), we have

$$\sum_i m_i b_i^2 = 1 + 2\frac{\epsilon}{\delta} + \frac{\epsilon^2}{\delta^2} \tag{96}$$

for the first term in the numerator. The second term in the numerator is

$$\sum_i m_i \Lambda^*(\Lambda'(b_i)) = \frac{1}{\epsilon^2}(\Lambda^* \circ \Lambda')(\epsilon) + 2\frac{\epsilon}{\delta^3}(\Lambda^* \circ \Lambda')(-\delta) + \frac{\epsilon^4}{\delta^6}(\Lambda^* \circ \Lambda')(\delta^2/\epsilon).$$

According to Lemma 27 and our assumptions on Λ, the last term is $o(\epsilon^2)$. Applying Lemma 28 to the other terms, we have

$$\begin{aligned}\sum_i m_i \Lambda^*(\Lambda'(b_i)) &= \frac{L}{2} + M\frac{\epsilon}{6} + L\frac{\epsilon}{\delta} - M\frac{\epsilon}{3} + O(\epsilon^2 + \epsilon\delta) \\ &= \frac{L}{2}\left(1 + \frac{2\epsilon}{\delta}\right) - M\frac{\epsilon}{6} + O(\epsilon^2 + \epsilon\delta),\end{aligned}$$

where

$$M = (\Lambda^*)'''(0)L^3 + 3\Lambda'''(0). \tag{97}$$

For the denominator in (85), we ignore the $i = 3$ contribution, giving a lower bound of

$$\begin{aligned}\sum_i m_i \Lambda'(b_i) &\geq \frac{\Lambda'(\epsilon)}{\epsilon} - 2\frac{\epsilon\Lambda'(-\delta)}{\delta^2} \\ &= \Lambda''(0) + \frac{\epsilon}{2}\Lambda'''(0) + O(\epsilon^2) + 2\frac{\epsilon}{\delta}\Lambda''(0) - \epsilon\Lambda'''(0) + O(\epsilon\delta) \\ &= L\left(1 + 2\frac{\epsilon}{\delta}\right) - \frac{\epsilon}{2}\Lambda'''(0) + O(\epsilon^2 + \epsilon\delta).\end{aligned}$$

Putting everything together,

$$\begin{aligned}&\frac{\sum_i m_i\Lambda'(b_i)\cdot\sum_i m_i\Lambda^*(\Lambda'(b_i))}{\left(\sum_i m_i\Lambda'(m_i)\right)^2} \\ &= \frac{\left(1 + \frac{2\epsilon}{\delta} + O(\epsilon^2)\right)\left(\frac{L}{2}\left(1 + \frac{2\epsilon}{\delta}\right) - \frac{\epsilon M}{6} + O(\epsilon^2 + \epsilon\delta)\right)}{\left(L\left(1 + \frac{2\epsilon}{\delta}\right) - \frac{\epsilon\Lambda'''(0)}{2} + O(\epsilon^2 + \epsilon\delta)\right)^2} \\ &= \frac{\frac{L}{2} - \frac{\epsilon M}{6} + O(\epsilon^2 + \epsilon\delta)}{L^2 - \epsilon L\Lambda'''(0) + O(\epsilon^2 + \epsilon\delta)} \\ &= \frac{1}{2L} - \frac{\epsilon M}{6L^2} + \frac{\epsilon\Lambda'''(0)}{2L^2} + O(\epsilon^2 + \epsilon\delta) \\ &= \frac{1}{2L} - \frac{\epsilon(\Lambda^*)'''(0)L}{6} + O(\epsilon^2 + \epsilon\delta),\end{aligned}$$

and in particular it is possible to choose δ and ϵ so that this quantity is at most $(1-\eta)\frac{1}{2L}$ for some $\eta > 0$.

Going back to (85) and recalling that the sequence a_i can be realized as the elements of a rank-1 p.s.d. matrix, M say, with $\|M\|_F = 1 + o(1)$, we have shown that

$$\ln\Pr(\lambda_1(A_n) > t) \geq \ln\Pr(\langle A, M\rangle > t\|M\|_F) \geq -(1-\eta)\frac{t^2}{4L} - o(t^2). \tag{98}$$

Replacing t by $m_n t$ and recalling that $L = \Lambda''(0) = \mathbb{E}\xi^2$ completes the proof of Theorem 9.

8. Back to triangle counts

Our eigenvalue LDP (Theorem 1) allows us to control the triangle-count contribution from a constant number of very extreme eigenvalues, but in order to fully characterize the behavior of the triangle-count, we need to handle the other eigenvalues also. We will do this in two steps: we use Theorem 11 to control the contribution of the bulk eigenvalues, and then Corollary 22 to show that the triangle count cannot be determined by $\omega(1)$ largish eigenvalues. Bear in mind that we will be applying our eigenvalue LDP to $\mathbb{E}A - A$, where A is the adjacency matrix, because Theorem 10 is for the positive eigenvalues of centered matrices, and we are interested in the negative eigenvalues here.

8.1. **The contribution of the bulk.** We consider two functions f_1 and f_2, where

$$f_1(x) = \begin{cases} 0 & \text{if } x < 0 \\ x^3 & \text{if } 0 \leq x < \sqrt{K} \\ 3Kx - 2K^{3/2} & \text{if } x \geq \sqrt{K} \end{cases} \tag{99}$$

and $f_2(x) = -f_1(-x)$. Then both f_1 and f_2 are $3K$-Lipschitz functions; also, f_1 is convex and f_2 is concave.

The following lemma is the main technical result of this section. Essentially, it says that changing the triangle-count using non-extreme eigenvalues carries a substantial entropy cost.

Lemma 29. *Let A_n be the centered adjacency matrix of a $\mathcal{G}(n,m)$ graph. There is a universal constant C such that if $K \geq C$ then*

$$\Pr\left(\sum_{i:\lambda_i(A_n)\geq -\sqrt{Kn}} \lambda_i^3(A_n) < -\delta - O(n^2)\right) \leq \exp\left(-\Omega\left(\frac{\delta^2}{n^3K^2}\right)\right). \tag{100}$$

Proof. We will prove the claim when A_n is the centered adjacency matrix of a $\mathcal{G}(n,p)$ graph, with $p = m/\binom{n}{2}$. The result for $\mathcal{G}(n,m)$ follows from the fact that a $\mathcal{G}(n,m)$ graph can be obtained by starting from $\mathcal{G}(n,p)$ and conditioning on the (probability $\Omega(1/n)$) event that there are exactly m edges.

Note that

$$f_1(x) + f_2(x) \leq \begin{cases} 0 & \text{if } x < -\sqrt{K} \\ x^3 & \text{if } x \geq -\sqrt{K}. \end{cases} \tag{101}$$

Hence,

$$\sum_i (f_1+f_2)(n^{-1/2}\lambda_i(A_n)) \leq n^{-3/2} \sum_{i:\lambda_i(A_n)\geq -\sqrt{Kn}} \lambda_i^3(A_n). \tag{102}$$

Since $-f_2$ is convex, Theorem 11 applies to both f_1 and f_2, giving

$$\Pr\left(\frac{1}{n}\operatorname{tr}[(f_1+f_2)(n^{-1/2}A_n)] \leq \frac{1}{n}\mathbb{E}\operatorname{tr}[(f_1+f_2)(n^{-1/2}A_n)] - s\right) \leq 2\exp(-\Omega(n^2s^2/K^2)) \tag{103}$$

whenever $s = \omega(K/n)$. From the inequality above, we also have

$$\Pr\left(\sum_{i:\lambda_i(A_n)\geq -\sqrt{Kn}} \lambda_i^3(A_n) \leq n^{3/2}\mathbb{E}\operatorname{tr}(f_1+f_2)(n^{-1/2}A_n) - s\right) \leq 2\exp\left(-\Omega\left(\frac{s^2}{K^2n^3}\right)\right). \tag{104}$$

It remains to control $\mathbb{E}\operatorname{tr}[(f_1+f_2)(n^{-1/2}A_n)]$; specifically, we want to show that $\mathbb{E}\operatorname{tr}(f_1+f_2)(n^{-1/2}A_n)$ is close to $n^{-3/2}\mathbb{E}\operatorname{tr}(A_n^3)$ (which is $O(\sqrt{n})$). But note that

$$|\operatorname{tr}[(f_1+f_2)(n^{-1/2}A_n) - n^{-3/2}A_n^3]| \\ \leq n^{-3/2}\sum_{i:|\lambda_i|>\sqrt{Kn}} |\lambda_i(A_n)|^3 \leq n^{-1/2}|s_1(A_n)|^3 1_{\{|s_1(A_n)|>\sqrt{Kn}\}},$$

where $s_1(A_n)$ is the largest singular value of A_n. But Proposition 21 implies that if K is sufficiently large then $\mathbb{E}[|s_1(A_n)|^3 1_{\{|s_1(A_n)|>\sqrt{Kn}\}}] \le \exp(-\Omega(\sqrt{n}))$. Hence,

$$\Pr\left(\sum_{i:\lambda_i(A_n)\ge -\sqrt{Kn}} \lambda_i^3(A_n) \le n^{3/2}\mathbb{E}\,\mathrm{tr}(A_n^3) - s - \exp(-\Omega(\sqrt{n}))\right) \le 2\exp\left(-\Omega\left(\frac{s^2}{K^2 n^3}\right)\right). \tag{105}$$

Finally, note that $\mathbb{E}\,\mathrm{tr}(A_n^3) = O(n^2)$. □

We will be interested in applying Lemma 29 when $\delta \gg n^{9/4}$. In this case, the $O(n^2)$ term becomes negligible and the probability bound is at most $\exp(-\omega(n^{3/2}))$.

8.2. **Many large negative eigenvalues.** There is one situation that we still need to handle: the possibility that there are $\omega(1)$ eigenvalues smaller than $-\Omega(\sqrt{n})$, and $\omega(1)$ of these eigenvalues contribute to the triangle count.

The first observation is that although Corollary 22 is written for a fixed *number* of singular values, it can be easily transferred to an inequality for singular values above a certain threshold.

Corollary 30. *With the notation of Corollary 22, if $s_i = s_i(A)$ are the singular values of A then*

$$\ln\Pr\left(\sqrt{\sum_{s_i>\sqrt{Kn}} s_i^2} \ge t\right) \le -\frac{t^2 L}{2} + O\left(\frac{t^2}{K}\ln K\right) \tag{106}$$

The same bound holds if A is the centered adjacency matrix of a $\mathcal{G}(n,m)$ graph.

Proof. Set $k = \lceil t^2/(Kn)\rceil$ and observe that if $s_1,\dots,s_k \ge \sqrt{Kn}$ then $\sum_{i=1}^k s_i^2 \ge t$. Hence, we either have

$$\sum_{s_i>\sqrt{Kn}} s_i^2 \le \sum_{i=1}^k s_i^2, \tag{107}$$

or else $\sum_{i=1}^k s_i^2 \ge t$. It follows that

$$\ln\Pr\left(\sqrt{\sum_{s_i>\sqrt{Kn}} s_i^2} \ge t\right) \le \ln\Pr\left(\sqrt{\sum_{i=1}^k s_i^2} \ge t\right), \tag{108}$$

and we conclude by applying Corollary 22 with our choice of k.

Finally, if A is the centered adjacency matrix of a $\mathcal{G}(n,m)$ graph then we use the same argument that was used to extend Corollary 22 to the $\mathcal{G}(n,m)$ case, namely that a $\mathcal{G}(n,m)$ graph can be obtained by conditioning a $\mathcal{G}(n,p)$ graph on an event of $\Omega(n^{-1})$ probability. □

8.3. **The upper bound in Theorem 1.** Let A be the adjacency matrix of a $\mathcal{G}(n,m)$ graph and recall that $\tau(A) = \frac{\mathrm{tr}[A^3]}{n(n-1)(n-2)} = \frac{\mathrm{tr}[A^3]}{n^3} + O(1/n)$. Let $\tilde{A} = A - \mathbb{E}A$; by Corollary 3,

$$\Pr(\tau(A) \le p^3 - t) = \Pr(\mathrm{tr}[A^3] \le n^3p^3 - n^3 t + O(n^2)) \le \Pr(\mathrm{tr}[\tilde{A}^3] \le -n^3 t + O(n^2)). \tag{109}$$

Writing out $\mathrm{tr}[\tilde{A}^3] = \sum_i \lambda_i^3(\tilde{A})$, choose $K = \omega(1)$ and $\epsilon = o(1)$ such that $K/\epsilon = o(n^{1/2}t^{2/3})$. Applying Lemma 29 gives

$$\Pr\left(\sum_{i:\lambda_i \ge -\sqrt{Kn}} \lambda_i^3(\tilde{A}) < -\epsilon t n^3\right) \le \exp\left(-\Omega\left(\frac{\epsilon^2 t^2 n^3}{K^2}\right)\right) = \exp(-\omega(n^2 t^{2/3})) \tag{110}$$

On the other hand, Jensen's inequality implies that

$$\left|\sum_{i:\lambda_i < -\sqrt{Kn}} \lambda_i^3\right| \le \left(\sum_{i:\lambda_i < -\sqrt{Kn}} \lambda_i^2\right)^{3/2} \le \left(\sum_{i:s_i > \sqrt{Kn}} s_i^2\right)^{3/2}, \tag{111}$$

where $\lambda_i = \lambda_i(\tilde{A})$ and $s_i = s_i(\tilde{A})$. By Corollary 30 (and taking into account the fact that $\epsilon = o(1)$ and $K = \omega(1)$),

$$\begin{aligned}\Pr\left(\sum_{i:\lambda_i < -\sqrt{Kn}} \lambda_i^3(\tilde{A}) < -(1-\epsilon)tn^3\right) &\le \Pr\left(\sqrt{\sum_{i:s_i > \sqrt{Kn}} s_i^2} > (1-\epsilon)^{1/3} t^{1/3} n\right) \\ &\le \exp\left(-\frac{L}{2} t^{2/3} n^2 + o(t^{2/3}n^2)\right).\end{aligned}$$

Combined with (110), this yields

$$\ln \Pr\left(\mathrm{tr}[\tilde{A}^3] \le -tn^3\right) \le -\frac{Lt^{2/3}n^2}{2}(1 + o(1)). \tag{112}$$

Now we apply (109), noting that $n^3 t = \omega(n^2)$, and so $n^3 t + O(n^2) = n^3 t(1 + o(1))$, to get

$$\ln \Pr(\tau(A) \le p^3 - t) \le -\frac{Lt^{2/3}n^2}{2}(1 + o(1)). \tag{113}$$

Corollary 31. *Conditioned on $\tau(A) \le p^3 - t$, $\sum_{i:\lambda_i \le -\Omega(\sqrt{n})} \lambda_i^3(\tilde{A}) \le -tn^3(1 - o(1))$ with high probability.*

The other piece of information we can extract from our proof is that the vertex degrees of a triangle-deficient graph are close to constant.

Corollary 32. *Conditioned on $\tau(A) \le p^3 - t$, if $d_1, \dots, d_n$ are the vertex degrees of the graph then with high probability*

$$\sum_i (d_i - pn)^2 = o(tn^3). \tag{114}$$

Proof. Since $\sum_i d_i = 2m = n^2 p + O(n)$, we have $\sum_i (d_i - pn)^2 = \sum_i d_i^2 - p^2 n^3 + O(n^2)$. Going back to the proof of Lemma 2, we have

$$\mathrm{tr}[\tilde{A}^3] \le \mathrm{tr}[A^3] - p^3 n^3 - 3p \sum_i (d_i - pn)^2 + O(n^2). \tag{115}$$

It follows that

$$\Pr(\tau(A) \le p^3 - t) \le \Pr\left(\mathrm{tr}[\tilde{A}^3] \le -tn^3 - 3p\sum_i (d_i - pn)^2 + O(n^2)\right). \tag{116}$$

Hence, if $tn^3 \gg n^2$ then

$$\begin{aligned}\ln \Pr\left(\tau(A) \le p^3 - t \text{ and } \sum_i (d_i - pn)^2 \ge \epsilon t n^3\right) \\ \le \ln \Pr\left(\mathrm{tr}[\tilde{A}^3] \le -tn^3(1 + \Omega(\epsilon))\right) \\ \le -\frac{Lt^{2/3}n^2}{2}(1 + \Omega(\epsilon)).\end{aligned}$$

In particular,

$$\Pr\left(\tau(A) \le p^3 - t \text{ and } \sum_i (d_i - pn)^2 \ge \epsilon t n^3\right) = o\left(\Pr\left(\tau(A) \le p^3 - t\right)\right), \tag{117}$$

and claim follows. □

8.4. **The lower bound in Theorem 1.** In showing the lower bound of Theorem 1, we need to apply Corollary 4 (instead of Corollary 3 as in the upper bound), and therefore we need to control $\mathrm{tr}[\tilde{A}^3]$ and $\sum_i d_i^2$ simultaneously. To do this, take v and ℓ as in Corollary 25 (applied with $t = t^{2/3}n^2$). Now, let $\xi_1, \dots, \xi_{\binom{n}{2}}$ be some ordering of the upper diagonal of $\tilde{A}$, ordered so that the first $\xi_1, \dots, \xi_{\binom{\ell}{2}}$ correspond to the upper diagonal of the upper-left $\ell \times \ell$ principal submatrix. Then $\langle \tilde{A}, vv^T \rangle = 2\sum_{i=1}^{\binom{\ell}{2}} v_1^2 \xi_i$, and so conditioning on $\langle \tilde{A}, vv^T \rangle < -t^{1/3}n$ is equivalent to conditioning on $\sum_{i=1}^q \xi_i < -\frac{t^{1/3}n}{2v_1^2}$ (where we have set $q = \binom{\ell}{2}$).

Let Ω be the event that $\langle \tilde{A}, vv^T \rangle \le -t^{1/3}n$. To prove the lower bound of Theorem 1, we show three properties:

(1) $\ln \Pr(\Omega) \ge -\frac{t^{2/3}n^2 L}{2}(1 + o(1))$.
(2) Conditioned on Ω, $\sum_i d_i^2 = n^3p^2 + O(n^2)$ with high probability.
(3) Conditioned on Ω, $\mathrm{tr}[\tilde{A}^3] \le -tn^3(1 - o(1))$ with high probability.

Once these three properties are shown, the lower bound of Theorem 1 will follow easily from Lemma 4, because by setting $D = O(n^2)$, the three properties above imply that

$$\Pr\left(\mathrm{tr}[\tilde{A}^3] \le -tn^3 + O(D) \text{ and } \sum_i d_i^2 \le n^3p^2 + O(D)\right) = -\frac{t^{2/3}n^2 L}{2}(1 + o(1)). \tag{118}$$

Note that the first property follows immediately from Corollary 25.

Lemma 33. *The second property is true: conditioned on Ω, $\sum_i d_i^2 = n^3p^2 + O(n^2)$ with high probability.*

Proof. Recall that $q = \binom{\ell}{2} = \Theta(t^{1/3}n)$, and it follows that conditioned on $\sum_{i=1}^q \xi_i < -\frac{t^{1/3}n}{2v_1^2}$ we have $\sum_{i=1}^q \xi_i = -(1+o(1))\frac{t^{1/3}n}{2v_1^2}$ with high probability. Now for any $s = (1+o(1))\frac{t^{1/3}n}{2v_1^2}$, if we condition on $\sum_{i=1}^q \xi_i = -s$ then the distribution of A is determined: there are $qp - s = p\binom{\ell}{2} - \Theta(t^{1/3}n)$ edges uniformly distributed without replacement among the first ℓ vertices, and $m - qp + s$ edges uniformly distributed in the rest of the graph. In particular, conditioned on this event each of the first ℓ vertices has expected degree

$$p(\ell-1) - \Theta(t^{1/3}n\ell^{-1}) + p(n-\ell) + \Theta(t^{1/3}) = p(n-1) - \Theta(t^{1/6}n^{1/2}) \tag{119}$$

where the first two terms on the left express the expected number of edges between our vertex and the other first-ℓ vertices, and the remaining two terms express the expected number of edges between our vertex and the other $n-\ell$ vertices. Similarly, the expected degree of the other $n-\ell$ vertices is

$$p(n-1) + \Theta(t^{1/3}). \tag{120}$$

These degrees are concentrated around their expectations (since the degrees of the last $n-\ell$ vertices has a hypergeometric distribution, and the degrees of the first ℓ vertices can be written as the sum of two independent hypergeometric variables), with variance $O(n)$. It follows that with high probability (still conditioned on $\sum_{i=1}^q \xi_i = -s$),

$$\begin{aligned}
\sum_i (d_i - pn)^2 &\le 2\sum_{i=1}^{\ell}\left((d_i - \mathbb{E}d_i)^2 + (\mathbb{E}d_i - pn)^2\right) + \sum_{i=\ell+1}^{n}(d_i - pn)^2 \\
&= O(n^2) + O(t^{1/3}n^2) + O(n^2) \\
&= O(n^2).
\end{aligned}$$

□

Lemma 34. *The third property is true: conditioned on Ω, $\operatorname{tr}[\tilde{A}^3] \le -tn^3(1-o(1))$ with high probability.*

Proof. Conditioned on Ω, we have $\lambda_n(\tilde{A}) \le -t^{1/3}n(1-o(1))$, and so it suffices to show that the contribution of the other eigenvalues is negligible. Note that the upper bound of Theorem 1, together with the fact that $\ln \Pr(\Omega) \ge -\frac{t^{2/3}n^2L}{2}(1-o(1))$, implies that conditioned on Ω, $\|\tilde{A}\| \le t^{1/3}n(1+o(1))$. Hence,

$$\begin{aligned}
\operatorname{tr}[(\tilde{A} + t^{1/3}nvv^T)^3] &= \operatorname{tr}[\tilde{A}^3] + t^{1/3}nv^T\tilde{A}^2v + t^{2/3}n^2v^T\tilde{A}v - tn^3 \\
&\ge \operatorname{tr}[\tilde{A}^3] - tn^3(1+o(1))
\end{aligned} \tag{121}$$

with high probability on Ω. Let $\tilde{B} = \tilde{A} + t^{1/3}nvv^T$. The basic idea is that $\tilde{B}$ contains off-diagonal entries that are almost independent and almost mean-zero, and therefore $\mathbb{E}\operatorname{tr}[\tilde{B}^3]$ can be easily computed. The most tedious part is just to check that the off-diagonal entries of $\tilde{B}$ have mean sufficiently small.

Recall from the previous proof that it suffices to prove the claim conditioned on $\sum_i \xi_i = s$ for any $s = (1+o(1))\frac{t^{1/3}n}{2v_1^2}$; call this event Ω_s. Recall that $\ell = v_1^{-2} + O(1)$, and so $qv_1^4 =$

$\binom{\ell}{2}v_1^4 = \frac{1}{2} + O(v_1^4) = \frac{1}{2} + O(t^{-2/3}n^{-2})$. On Ω_s, therefore, we have

$$\begin{aligned}\sum_{i=1}^{q} \xi_i + t^{1/3}nv_1^2 &= s + qt^{1/3}nv_1^2 \\ &= -(1+o(1))\frac{t^{1/3}n}{2v_1^2} + \frac{t^{1/3}n}{2v_1^2} + O(t^{-1/3}n^{-1}) \\ &= O(t^{-1/3}n^{-1}).\end{aligned}$$

In particular, if ω_i are the upper-diagonal entries of the matrix $\tilde{B}$ (ordered the same way as ξ_i), then

- for $1 \le i \le q$, $\mathbb{E}[\omega_i \mid \Omega_s] = O(q^{-1}t^{-1/3}n^{-1}) = O(t^{-2/3}n^{-2})$, and
- for $q < i \le \binom{n}{2}$, $\mathbb{E}[\omega_i \mid \Omega_s] = \mathbb{E}[\xi_i \mid \Omega_s] = O(t^{1/3}n^{-1})$.

Hence, all off-diagonal entries of $\tilde{B}$ have expectation (conditioned on Ω_s) $o(n^{-1})$.

Let b_{ij} be the entries of $\tilde{B}$. These are not independent given Ω_s; they are determined by sampling without replacement among the entries of the first $\ell \times \ell$ submatrix, and then again by sampling without replacement among the remaining entries. However, the b_{ij} are almost independent in the sense that $\mathbb{E}[b_{i_1,j_1}^{r_1} b_{i_2,j_2}^{r_2} b_{i_3,j_3}^{r_3}] = (1+o(1))\mathbb{E}[b_{i_1,j_1}^{r_1}]\mathbb{E}[b_{i_2,j_2}^{r_2}]\mathbb{E}[b_{i_3,j_3}^{r_3}]$ for any fixed choice of entries and exponents. Also, the diagonal entries of $\tilde{B}$ are uniformly bounded.

$$\mathbb{E}[\mathrm{tr}[\tilde{B}^3 \mid \Omega_s] = \sum_{i,j,k} \mathbb{E}[b_{i,j}b_{j,k}b_{k,i}] = O(n), \tag{122}$$

with the dominant contribution coming from the terms with $i = j = k$. By Markov's inequality and the fact that $n = o(tn^3)$, it follows that $\mathrm{tr}[\tilde{B}^3] = o(tn^3)$ with high probability conditioned on Ω_s. Going back to (121), the claim follows. □

8.5. **The two extreme eigenvalues.** In proving the upper bound on $\Pr(\tau(A) \le p^3 - t)$, we applied the inequality $\sum_i |a_i|^3 \le (\sum_i a_i^2)^{3/2}$ to the collection of most-negative eigenvalues. In order to understand how these most negative eigenvalues are actually distributed, observe that in order for the inequality above to be an equality, all but one of the terms in the sum must be zero. Made quantitative, this observation implies that in order for our probability upper bound to be tight, the smallest eigenvalue must dominate the others.

Lemma 35. *Let $a_1, \dots$ be a sequence of non-negative numbers, in non-increasing order. If*

$$\sum_{i \ge 2} a_i^3 \ge \epsilon \|a\|_3^3 \tag{123}$$

then

$$\|a\|_2^2 \ge (1+\epsilon)^{1/3} \|a\|_3^{2/3}. \tag{124}$$

Proof. For $c > b > 0$,

$$(b^2+c^2)^3 - (b^3+c^3)^2 = 3b^2c^2(b^2+c^2) - 2b^3c^3 \ge 4b^3c^3 \ge \frac{b^3}{c^3}(b^3+c^3)^2, \tag{125}$$

or in other words,

$$b^3 + c^3 \le \frac{1}{\left(1 + \frac{b^3}{c^3}\right)^{1/2}} (b^2 + c^2)^{3/2}. \tag{126}$$

Applying this with $c = a_1$ and $b = \left(\sum_{i \ge 2} a_i^3\right)^{1/3}$, note that our assumptions imply that $b^3/c^3 \ge \epsilon$, and so

$$\|a\|_3^3 \le \frac{1}{\sqrt{1+\epsilon}} \left(a_1^2 + \left(\sum_{i \ge 2} a_i^3 \right)^{2/3} \right)^{3/2} \le \frac{1}{\sqrt{1+\epsilon}} \left(\sum_{i \ge 1} a_i^2 \right)^{3/2}, \tag{127}$$

where the final inequality follows from Jensen's inequality. □

Applying Lemma 35 to the most negative eigenvalues of $\tilde{A}$ allows us to show that the eigenvalues of $\tilde{A}$ satisfy the claims that Theorem 1 makes for the eigenvalues of A.

Corollary 36. *In the setting of Theorem 1, for any $\epsilon > 0$, conditioned on $\tau(A) \le p^3 - t$ we have*

$$\lambda_n^3(\tilde{A}) \le -(1-\epsilon) t n^3 \text{ and } \lambda_{n-1}^3(\tilde{A}) \ge -\epsilon t n^3 \tag{128}$$

with high probability.

Proof. Let $S = \{i : \lambda_i(\tilde{A}) \le -\Omega(\sqrt{n})\}$. By Corollary 31, for any $\delta > 0$, conditioned on $\tau(A) \le p^3 - t$ we have

$$\sum_{i \in S} \lambda_i^3(\tilde{A}) \le -(1-\delta) t n^3 \tag{129}$$

with high probability. On this event, we either have $\lambda_n^3(\tilde{A}) \le -(1-\delta-\epsilon) t n^3$ or $\sum_{i \in S \setminus \{n\}} \lambda_i^3(\tilde{A}) \le -\epsilon t n^3$. We will show that for some $\delta = \Omega(\epsilon)$,

$$\Pr\left(\sum_{i \in S} \lambda_i^3(\tilde{A}) \le -(1-\delta) t n^3 \text{ and } \sum_{i \in S \setminus \{n\}} \lambda_i^3(\tilde{A}) \le -\epsilon t n^3 \right) \tag{130}$$

is much smaller than $\Pr(\tau(A) \le p^3 - t)$; this will imply the claim.

Indeed, applying Lemma 35 to the sequence of $|\lambda_i|$ for $i \in S$, we see that if

$$\sum_{i \in S} \lambda_i^3(\tilde{A}) \le -(1-\delta) t n^3 \text{ and } \sum_{i \in S \setminus \{n\}} \lambda_i^3(\tilde{A}) \le -\epsilon t n^3 \tag{131}$$

then

$$\sum_{i \in S} \lambda_i^2(\tilde{A}) \ge (1+\epsilon)^{1/3} (1-\delta) t^{2/3} n^2 \ge (1 + \Omega(\epsilon)) t^{2/3} n^2, \tag{132}$$

where the last inequality follows by choosing a sufficiently small $\delta = \Omega(\epsilon)$. But Corollary 30 implies that

$$\begin{aligned} \Pr\left(\sum_{i \in S} \lambda_i^2(\tilde{A}) \ge (1 + \Omega(\epsilon)) t^{2/3} n^2 \right) &\le \exp\left(-(1 + \Omega(\epsilon))(1 - o(1)) \frac{t^{2/3} n^2 L}{2} \right) \\ &= o(\Pr(\tau(A) \le p^3 - t)), \end{aligned}$$

where the final bound follows from the lower bound of Theorem 1. □

To complete the proof of Theorem 1, we need to pass from the eigenvalues of $\tilde{A}$ to the eigenvalues of A, recall that $A = \tilde{A} + p\mathbf{1} - pI$. Since $p\mathbf{1} \geq 0$, we have

$$\lambda_{n-1}(A) \geq \lambda_{n-1}(\tilde{A}) - p, \tag{133}$$

and so $\lambda_{n-1}(\tilde{A}) \geq -o(tn^3)$ implies the same for $\lambda_{n-1}(A)$. For λ_n, we have $\lambda_n(A) \leq \lambda_n(\tilde{A}+p\mathbf{1})$ and so it remains to take care of the $p\mathbf{1}$ term.

Let v be an eigenvector with eigenvalue $\lambda_n(\tilde{A})$ satisfying $|v|^2 = 1$. Then $|\langle v, \tilde{A}\mathbf{1}\rangle| = |\lambda_n(\tilde{A})||\langle v, \mathbf{1}\rangle|$; conditioned on $\tau(A) \leq p^3 - t$, this is $(1+o(1))tn^3|\langle v, \mathbf{1}\rangle|$ with high probability. On the other hand $\tilde{A}\mathbf{1} = (A - p\mathbf{1} + pI)\mathbf{1} = d - p(n-1)\mathbf{1}$, where d whose entries are the vertex degrees. By Corollary 32, conditioned on $\tau(A) \leq p^3 - t$ we have $|d - pn\mathbf{1}|^2 = o(tn^3)$, and since $|\mathbf{1}|^2 = n = o(tn^3)$, we also have $|\tilde{A}\mathbf{1}|^2 = |d - p(n-1)\mathbf{1}|^2 = o(tn^3)$. Hence,

$$(1 + o(1))tn^3|\langle v, \mathbf{1}\rangle| = |\langle v, \tilde{A}\mathbf{1}\rangle| \leq |\tilde{A}\mathbf{1}| = o(tn^3), \tag{134}$$

and it follows that $|\langle v, \mathbf{1}\rangle| = o(1)$. Finally, note that

$$v^T(\tilde{A} + p\mathbf{1})v = v^T\tilde{A}v + p|\langle v, \mathbf{1}\rangle|^2 = \lambda_n(\tilde{A}) + o(1), \tag{135}$$

and so by Rayleigh's criterion it follows that $\lambda_n(\tilde{A}+p\mathbf{1}) \leq \lambda_n(\tilde{A}) + o(1)$. This completes the proof of Theorem 1.

References

[1] Christina Goldschmidt, Simon Griffiths and Alex Scott, Moderate deviations of subgraph counts in the Erdős-Rényi random graphs $G(n,m)$ and $G(n,p)$, Trans. Amer. Math. Soc. **373** (2020) 5517–5585.

[2] Svante Janson, Poisson approximation for large deviations, Random Structures & Algorithms **1** (1990) 221–229.

[3] Stéphane Boucheron, Gábor Lugosi, Pascal Massart etal, Concentration inequalities using the entropy method, Ann. Prob. **31** (2013) 2003 1583–1614.

[4] Sourav Chatterjee, The missing log in large deviations for triangle counts, Random Structures & Algorithms **40** (2012) 437–451.

[5] S. Janson and A. Ruciński, The Infamous Upper Tail, Random Structures & Algorithms **20** (2002 317-342.

[6] C. Radin and L. Sadun, Phase transitions in a complex network, J. Phys. A **46** (2013) 12 pp.

[7] C. Radin and L. Sadun, Singularities in the entropy of asymptotically large simple graphs, J. Stat. Phys. **158** (2015) 853–865.

[8] Svante Janson, *Orthogonal decompositions and functional limit theorems for random graph statistics*, American Mathematical Soc. 1994.

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

[10] J. Neeman, C. Radin and L. Sadun, Phase transitions in finite random networks, J. Stat. Phys. **181** (2020) 30–328.

[11] S. Chatterjee and S.R.S. Varadhan, The large deviation principle for the Erdős-Rényi random graph, *Eur. J. Combin.* **32** (2011) 1000-1017.

[12] A. Dembo and E. Lubetzky, A large deviation principle for the Erdős-Rényi uniform graph, Electron. Commun. Probab. **23** 1-13.

[13] Alice Guionnet and Jonathan Husson, Large deviations for the largest eigenvalue of Rademacher matrices, Ann. Prob. **48** (2020) 1436–1465.

[14] Fanny Augeri, Large deviations principle for the largest eigenvalue of Wigner matrices without Gaussian tail, Electron. J. Probab. **21** (2016) 49 pp.

[15] Bhaswar B. Bhattacharya and Shirshendu Ganguly, Upper Tails for Edge Eigenvalues of Random Graphs, SIAM Journal on Discrete Mathematics **34** (2020) 1069–1083.
[16] Fanny Augeri, Alice Guionnet and Jonathan Husson, Large deviations for the largest eigenvalue of sub-Gaussian matrices, arXiv:1911.10591.
[17] Alice Guionnet and Ofer Zeitouni, Concentration of the Spectral Measure for Large Matrices, Electron. Commun. Probab. **5** (2020) 119–136.
[18] Terence Tao, *Topics in random matrix theory*, Americal Mathematical Soc. 1012
[19] Roman Vershynin, *High-dimensional probability: An introduction with applications in data science*, Cambridge University Press 2018.
[20] Olav Kallenberg, Foundations of Modern Probability, Second Edition, Springer (2001)

Joe Neeman, Department of Mathematics, The University of Texas at Austin, Austin, TX 78712

Email address: joeneeman@gmail.com

Charles Radin, Department of Mathematics, The University of Texas at Austin, Austin, TX 78712

Email address: radin@math.utexas.edu

Lorenzo Sadun, Department of Mathematics, The University of Texas at Austin, Austin, TX 78712

Email address: sadun@math.utexas.edu