MODERATE DEVIATIONS IN CYCLE COUNT

JOE NEEMAN, CHARLES RADIN, AND LORENZO SADUN

ABSTRACT. We prove moderate deviations bounds for the lower tail of the number of odd cycles in a $\mathcal{G}(n,m)$ random graph. We show that the probability of decreasing triangle density by t^3, is $\exp(-\Theta(n^2t^2))$ whenever $n^{-3/4} \ll t^3 \ll 1$, while for $k \geq 5$ we give the same estimate for the probability of decreasing the k-cycle density by t^k, but for the larger range $n^{-1} \ll t^k \ll 1$. When $m \geq \frac{1}{2}\binom{n}{2}$, we also find the leading coefficient in the exponent. This complements results of Goldschmidt et al., who showed that for $n^{-3/2} \ll t^k \ll n^{-1}$, the probability is $\exp(-\Theta(n^3t^{2k}))$. That is, deviations of order smaller than n^{-1} behave like small deviations, and deviations of order larger than $n^{-3/4}$ (for triangles) or n^{-1} (for k-cycles with $k \geq 5$) behave like large deviations. For triangles, we conjecture that a sharp change between the two regimes occurs for deviations of size $n^{-3/4}$, which we associate with a single large negative eigenvalue of the adjacency matrix becoming responsible for almost all of the cycle deficit.

Our results can be interpreted as finite size effects in phase transitions in constrained random graphs.

1. Introduction

We prove moderate deviations bounds for the lower tail of the number of odd k-cycles in a $\mathcal{G}(n,m)$ random graph, i.e. a uniformly random graph among all the graphs with n vertices and m edges. We study deviations larger than those of Goldschmidt et al. [1] but smaller than large deviations, which are of order the mean of the cycle density. For instance, with the notation that $\tau_3(G)$ 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 $1/2 \leq p < 1$ that is fixed as $n \to \infty$ and $n^{-3/4} \ll t^3 \ll 1$, we prove (see Theorem 1) that

$$\Pr\left(\tau_3(G) \leq p^3 - t^3\right) = \exp\left(-\frac{\ln\frac{p}{1-p}}{2(2p-1)}t^2n^2 + o(t^2n^2)\right). \tag{1}$$

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].

Date: January 7, 2022.

This work was partially supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2047/1 – 390685813, and by a fellowship from the Alfred P. Sloan Foundation.

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 for $p \to 0$, and a substantial part of the difficulty in the study of the upper tail is to obtain the correct dependence on p. For dense graphs (i.e. when p is fixed), the lower tail has more subtle behavior, as was noted already by [12]. In this regime the $\mathcal{G}(n,m)$ model, in which the number of edges is fixed at m, differs substantially from the $\mathcal{G}(n,p)$ model. For example, one can easily see that under $\mathcal{G}(n,p)$, the number of triangles, $T_3(G)$, satisfies $\mathrm{Var}(T_3(G)) = \Theta(n^4)$, while under $\mathcal{G}(n,m)$, $\mathrm{Var}(T_3(G)) = \Theta(n^3)$. 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 $\tau_3 \approx q^3$, and the probability of seeing such 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

$$\Pr(\tau_3(G) \le \mathbb{E}\tau_3(G) - t^3) \ge \exp(-\Omega(n^2 t^6)). \tag{2}$$

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 $m = \Theta(n^2)$,

$$\Pr(\tau_3(G) \le \mathbb{E}\tau_3(G) - t^3) = \exp(-\Theta(n^2 t^2)). \tag{3}$$

(They also found the exact leading-order term in the exponent when $m = \frac{1}{2}\binom{n}{2} + o(n^2)$ and bounded the leading-order coefficient for all other values of m.) The same argument also works for odd $k > 3$. At the other end of the scale, a recent result of Goldschmidt et al. [1] showed that for $n^{-3/2} \ll t^k \ll n^{-1}$ the lower tail has a different behavior:

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

(Again, they also found the exact leading-order term in the exponent.) Since $t^k \le \Theta(n^{-3/2})$ is within the range of the Central Limit Theorem this leaves open the case of $n^{-1} \ll t^k \ll 1$. Noting that the two exponential rates (namely n^2t^2 and n^3t^{2k}) cross over at $t^k = \Theta(n^{-\frac{k}{2(k-1)}})$, it is natural to guess that for all odd k,

$$\Pr(\tau_k(G) \le \mathbb{E}\tau_k(G) - t^k) = \begin{cases} \exp(-\Theta(n^3 t^{2k})) & \text{if } t^k \ll n^{-\frac{k}{2(k-1)}}, \\ \exp(-\Theta(n^2 t^2)) & \text{if } n^{-\frac{k}{2(k-1)}} \ll t^k \ll 1. \end{cases} \tag{5}$$

In the case $k = 3$, we prove the second of these two cases; the first remains a conjecture. For $k \ge 5$, (5) turns out to be false: the boundary between the two regimes turns out to occur when t^k is of the order n^{-1}. This is perhaps surprising because it implies that a deviation of order $n^{-1-\epsilon}$ has probability $\exp(-\Theta(n^{1+2\epsilon}))$ but a deviation of order $n^{-1+\epsilon}$ has the much smaller probability $\exp(-n^{2-2/k-O(\epsilon)})$.

We also prove some structural results on graphs with $\tau_k(G) \le \mathbb{E}\tau_k(G) - t^k$ in our range of t^k: conditioned on this cycle-count deviation, with high probability such a graph has a very negative eigenvalue, and also has a small subgraph with substantially smaller edge density. These structural results provide a plausible explanation for the importance of the threshold between the two regimes: it is the threshold at which a single large negative eigenvalue of the adjacency matrix becomes responsible for almost all of the k-cycle deficit.

2. Context and references

We are concerned with random graphs $\mathcal{G}(n,m)$, the uniform distribution on graphs on n nodes with m edges. For a graph G and an integer $k \geq 3$, define $T_k(G)$ to be the number of injective maps $\phi : \{1,\dots,k\} \to V(G)$ for which $\{\phi(1),\phi(2)\},\{\phi(2),\phi(3)\},\dots,\{\phi(k),\phi(1)\}$ are all edges of G; we say that $T_k(G)$ is the number of k-cycles in G. The *k-cycle density* is $\tau_k(G) = \frac{1}{\binom{n}{k}} T_k(G) \in [0,1]$. Results on the probability of deviations of subgraph density from the mean fall into three classes by size: small deviations, on the order of the standard deviation, large deviations, on the order of the mean, and moderate deviations, of intermediate size.

Our main results concern the moderate regime of deviations of cycle density in $\mathcal{G}(n,m)$, in which we prove, among other things, that deviations near but below the large class are qualitatively different from deviations near but above the small class. We know of no other results of this sort, for $\mathcal{G}(n,m)$ or the $\mathcal{G}(n,p)$ random graph model, in which edges appear independently.

For small deviations there is a long history under the name Central Limit Theorem. There are also many papers on moderate and large deviations of subgraph counts. As background, more specifically for results discussed here, we suggest the following: [8, 9, 10, 11, 12, 13, 14, 15, 16] and references within them for a broader view. As our results are strongly colored by large deviations we note in particular [17].

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 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 k-cycle density of $\mathcal{G}(n,m)$ for $t \to 0$ (but not too quickly) and for odd k (for even k, it is not possible for $\tau_k(G)$ to be significantly smaller than p^k). The case $0 \leq t^k \leq \Omega(n^{-3/2})$ is within the range of the Central Limit Theorem and it is covered by Janson's more general work on subgraph statistics [18]. The range $n^{-3/2} \ll t^k \ll n^{-1}$ is studied by [1]; they showed that in this regime

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

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

$$\Pr(\tau_k(G) \leq \mathbb{E}\tau_k(G) - t^k) = \exp\left(-\Omega(t^k n^2)\right). \tag{7}$$

We show that this upper bound is mostly not tight. In particular, for k-cycles with $k \geq 5$ we show that the correct exponent is t^2n^2 for all $n^{-1} \ll t^k \ll 1$. For triangles, we show the same exponent but only in the range $n^{-3/4} \ll t^3 \ll 1$; we conjecture that this is the best possible range, and that the bound (6) is sharp for triangles in the range $n^{-1} \ll t^3 \ll n^{-3/4}$. In the case $p \geq \frac{1}{2}$, we also derive more detailed results (see Theorem 1): we identify the leading constant in the exponent and we prove some results on the graph structure conditioned on having few cycles.

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 [19, 9]. In particular, [19] is the only existing work we know of in which the conditional structure of subgraph-density-constrained random graphs is established. Specifically, they show that for sparse random graphs conditioned on having more than the expected number of cliques, the random graph has either a "clique" structure in which there is a collection of vertices has higher-than-expected edge density or a "hub" structure in which there is a partition of the vertices with a higher-than-expected edge density between the two parts. In contrast, our results show that for dense random graphs with fewer cycles than expected, there is a collection of vertices with lower-than-expected edge density.

Moderate deviations in triangle count (i.e. the case $k = 3$) in $\mathcal{G}(n, m)$ can be seen from a different vantage based on [20]. That paper follows a series of works [6, 7, 24, 22, 23, 25, 26] on the asymptotics of 'constrained' random graphs, in particular the asymptotics of $\mathcal{G}(n, m, t)$, the uniform distribution on 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 [27] by Chatterjee and Varadhan on large deviations in $\mathcal{G}(n, p)$, was used to prove various features of phase transitions between asymptotic 'phases', phases illustrated by the entropy-optimal graphons. (See also [28].) But in [20] 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 $2\sigma\sqrt{n}$. However, because an eigenvalue of order n^β contributes $n^{k\beta}$ to the k-cycle count, and because we are interested in cycle-count deviation of orders larger than $n^{k/2}$, we are necessarily 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 cycle count is affected by several atypically large eigenvalues instead of just one.

In related works,

- Guionnet and Husson [29] 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 [30] showed an LDP for the largest eigenvalue for random matrices whose entries have heavier-than-Gaussian tails.
- Battacharya and Ganguly [31] showed an LDP for the largest two eigenvalues of a sparse Erdős-Rényi graph. The methods we use for our eigenvalue LDPs are related to their methods for the second-largest eigenvalue. In order to make the connection to cycle counts, however, we need to handle the entire spectrum.
- Augeri, Guionnet, and Husson [32] showed an LDP for the largest eigenvalue for most random matrices with subgaussian elements. These are 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. Cycle 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 $\tau_k(A)$ the k-cycle density of A, and we order the eigenvalues $\lambda_1(A) \geq \cdots \geq \lambda_n(A)$ in non-increasing order.

We prove two theorems governing asymptotic behavior as $n \to \infty$. We define the critical exponent

$$c_* = \min\{1, \frac{k(2-k)}{2k-2}\} \tag{8}$$

and we assume that $n^{-c_*} \ll t^k \ll 1$; this is equivalent to $n^{-3/4} \ll t^3 \ll 1$ for $k = 3$, and $n^{-1} \ll t^k \ll 1$ for $k \geq 5$. Our first theorem is a strong result for $\frac{1}{2} \leq p < 1$.

Theorem 1. *If $\frac{1}{2} \leq p < 1$ and $n^{-c_*} \ll t^k \ll 1$ then*

$$\Pr\left(\tau_k(A) \leq p^k - t^k\right) = \exp\left(-\frac{\ln\frac{1-p}{p}}{2(1-2p)} t^2 n^2 + o(t^2 n^2)\right), \tag{9}$$

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

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

and $\lambda_{n-1}(A) \geq -o(tn)$.

The second result, for $0 < p \leq \frac{1}{2}$, is weaker.

Theorem 2. *If $0 < p \leq \frac{1}{2}$ and $n^{-c_*} \ll t^k \ll 1$ then $\Pr\left(\tau_k(G) \leq p^k - t^k\right)$ is bounded above by*

$$\exp\left(-\frac{\ln\frac{p}{1-p}}{2(2p-1)} t^2 n^2 + o(t^2 n^2)\right) \tag{11}$$

and bounded below by

$$\exp\left(-\frac{1}{2p(1-p)}t^2n^2 + o(t^2n^2)\right). \tag{12}$$

Moreover, conditioned on $\tau_k(A) \le p^k - t^k$, with high probability we have

$$\lambda_n(A) = -\Omega(tn). \tag{13}$$

Together, these theorems show that $\Pr(\tau_k(A) \le p^k - t^k) = \exp(-\Theta(t^2n^2))$ for all $0 < p < 1$ and $n^{-c_*} \ll t^k \ll 1$.

In the case $p \ge \frac{1}{2}$, we also give a graph-theoretic characterization of the conditioned graph: given that $\tau_k(A) \le p^k - t^k$, the graph contains a lower-density subgraph of about $tn/(2p-1)$ vertices. In what follows, for $V_1, V_2 \subset V$, let

$$E(V_1, V_2) = \sum_{v_1 \in V_1, v_2 \in V_2} 1_{\{\{v_1,v_2\} \in E(G)\}}$$

count the edges between V_1 and V_2, while double-counting those edges with both endpoints in $V_1 \cap V_2$.

Theorem 3. *If $\frac{1}{2} \le p < 1$ and $n^{-k/(2(k-1))} \ll t \ll 1$ then conditioned on $\tau_k(G) \le p^k - t^k$, with high probability there exists a subset $U \subset V(G)$ of size $|U| = (1 + o(1))tn/(2p-1)$ such that for every $V_1, V_2 \subset V(G)$,*

$$E(V_1, V_2) = p|V_1||V_2| - (2p-1)|V_1 \cap U||V_2 \cap U| + o(tn|V_1 \cup V_2|).$$

In particular, setting $V_1 = V_2 = U$ shows that the subgraph induced by U has edge density about $1 - p$. More generally, Theorem 3 implies that G has no other non-trivial structure at the scale of tn or more vertices.

3.1. **Centering the matrix.** The main point of this section is that when considering the lower tail for cycle 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 – as discussed in the introduction – in $\mathcal{G}(n, p)$ the entropically favorable way to reduce the k-cycle 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 4. *Let A be the adjacency matrix of a graph with n vertices and $m = p\binom{n}{2}$ edges, and let d_i be the degree of vertex i. Let $\tilde{A} = A - p\mathbf{1} + pI$. For any $k \ge 3$, there exists $\epsilon > 0$ such that if $\|\tilde{A}\|_{op} \le \epsilon n$ then*

$$\mathrm{tr}[\tilde{A}^k] = \mathrm{tr}[A^k] - p^k n^k - (1 - O(\epsilon))kn^{k-3} \sum_i (d_i - pn)^2 + O(n^{k-1}) \tag{14}$$

Proof. Let $B = \tilde{A} + p\mathbf{1}$ and consider $\operatorname{tr}[B^k]$. (The extra contribution of pI in A makes a lower-order contribution and we will handle it later.) Consider the various terms in the expansion $(\tilde{A} + p\mathbf{1})^k$ according to how many copies of $\tilde{A}$ they contain: there is a $\tilde{A}^k$ term and a $p^k n^{k-1}\mathbf{1}$ term (which has trace $p^k n^k$), and every other term is a product involving at least one occurence of $\mathbf{1}$ and at least one occurrence of $\tilde{A}$. Note that $\mathbf{1}\tilde{A}\mathbf{1} = 0$, and so all the terms that have exactly one occurrence of $\tilde{A}$ vanish; and of the terms containing exactly two occurrences of $\tilde{A}$, the only non-vanishing ones are of the form $\tilde{A}^2\mathbf{1}_{k-2}$ (up to cyclic permutation). There are k of these terms, and so after taking the trace, they contribute

$$k \operatorname{tr}[\tilde{A}^2\mathbf{1}^{k-2}] = kn^{k-3}\operatorname{tr}[\tilde{A}^2\mathbf{1}] = kn^{k-3}|\tilde{A}1|^2 = kn^{k-3}\sum_i (d_i - pn)^2 \tag{15}$$

to $\operatorname{tr}[B^k]$.

Next, consider the terms containing more than two occurrences of $\tilde{A}$. Since $\mathbf{1}\tilde{A}\mathbf{1} = 0$, the only non-vanishing contributions take the form

$$\operatorname{tr}\prod_{i=1}^m \mathbf{1}^{j_i}\tilde{A}^{\ell_i}$$

for some $\ell_i \geq 2$, and $\sum_i \ell_i \geq 3$. Since $\mathbf{1}^j = n^{j-1}\mathbf{1}$, the term displayed above can be re-written (setting $j = \sum j_i = k - \sum \ell_i$) as

$$n^{j-m}\operatorname{tr}\prod_{i=1}^m \mathbf{1}\tilde{A}^{\ell_i} = n^{j-m}\prod_{i=1}^m 1^T\tilde{A}^{\ell_i}1.$$

Since each $\ell_i \geq 2$, if $\|\tilde{A}\|_{\mathrm{op}} \leq \epsilon n$ then $|1^T\tilde{A}^{\ell_i}1| \leq \|\tilde{A}^{\ell_i-2}\|_{\mathrm{op}}|\tilde{A}1|^2 \leq \epsilon^{\ell_i-2}n^{\ell_i-2}|\tilde{A}1|^2 \leq \epsilon^{\ell_i}n^{\ell_i+1}$. Now we consider two cases: if $\ell_i = 2$ for all i then $m \geq 2$ (because $\sum_i \ell_i \geq 3$). In this case, we use the bound $|1^T\tilde{A}^{\ell_i}1| \leq \epsilon^{\ell_i}n^{\ell_i+1} \leq \epsilon^2 n^{\ell_i+1}$ for $i \geq 2$ and the bound $|1^T\tilde{A}^{\ell_i}1| \leq \epsilon^{\ell_i-2}n^{\ell_i-2}|\tilde{A}1|^2 \leq n^{\ell_i-2}|\tilde{A}1|^2$ for $i = 1$, to obtain

$$\left|\operatorname{tr}\prod_{i=1}^m \mathbf{1}^{j_i}\tilde{A}^{\ell_i}\right| \leq \epsilon^2 n^{k-3}|\tilde{A}1|^2. \tag{16}$$

(The $k-3$ exponent on n comes from the fact that $j - m + \ell_1 - 2 + \sum_{i=2}^m(\ell_i + 1) = j + \sum_i \ell_i - 3 = k - 3$.) On the other hand, if there is some i with $\ell_i \geq 3$ then without loss of generality $i = 1$; we apply the bound $|1^T\tilde{A}^{\ell_i}1| \leq \epsilon^{\ell_i}n^{\ell_i+1} \leq n^{\ell_i+1}$ for $i \geq 2$ and the bound $|1^T\tilde{A}^{\ell_i}1| \leq \epsilon^{\ell_i-2}n^{\ell_i-2}|\tilde{A}1|^2 \leq \epsilon n^{\ell_i-2}|\tilde{A}1|^2$ for $i = 1$, to obtain

$$\left|\operatorname{tr}\prod_{i=1}^m \mathbf{1}^{j_i}\tilde{A}^{\ell_i}\right| \leq \epsilon n^{k-3}|\tilde{A}1|^2. \tag{17}$$

Now compare (15) to (16) and (17): out of all the terms in the expansion of $\operatorname{tr}[(\tilde{A} + p\mathbf{1})^k]$ that contain between 1 and $k-1$ copies of $\tilde{A}$, the terms containing two adjacent copies of $\tilde{A}$ (i.e. the terms we compute in (15) dominate). Since the total number of terms in the expansion is 2^k, we see that if ϵ is sufficiently small in terms of k then

$$\operatorname{tr}[B^k] = \operatorname{tr}[\tilde{A}^k] + p^k n^k + (1 - O(\epsilon))kn^{k-3}\sum_i (d_i - pn)^2.$$

Finally, to get the claim in terms of $A = B - pI$, note that $A^k = \sum_{j=0}^{k} \binom{k}{j} B^j (-p)^j$. We apply our previous result to each B^j term, noting that for $j \geq 1$ each term contributes only $O(n^{k-1})$. □

Combining Lemma 4 with the observation that $\mathbb{E}\,\mathrm{tr}[A^k] = p^k n^k + O(n^{k-1})$ when A is the adjacency matrix of a $\mathcal{G}(n, m)$ graph, we arrive at the following consequence:

Corollary 5. *Let A be the adjacency matrix of a $\mathcal{G}(n, m)$ graph and let $\tilde{A} = A - \mathbb{E}A$. For any $t \geq 0$ and all sufficiently small $\epsilon > 0$ depending on k,*

$$\Pr(\mathrm{tr}[A^k] \leq \mathbb{E}\,\mathrm{tr}[A^k] - t^k) \leq \Pr(\mathrm{tr}[\tilde{A}^k] \leq -t^k + O(n^{k-1})) + \Pr(\|\tilde{A}\|_{op} \geq \epsilon n) \tag{18}$$

4. Large deviations for eigenvalues of random matrices

In this section and beyond, we let A denote a generic random matrix and we estimate the most positive eigenvalues of A. Since we are looking at lower tails, the most important such matrix to keep in mind is *minus* the centered adjacency matrix, previously denoted $\tilde{A}$ or $A - \mathbb{E}A$. This is the same as *plus* the centered adjacency matrix of a random graph with edge density $q = 1 - p$. The proof of Theorem 1 ($p \geq \frac{1}{2}$) thus relies on results for $q \leq \frac{1}{2}$, while the proof of Theorem 2 ($p \leq \frac{1}{2}$) relies on results for $q \geq \frac{1}{2}$.

Definition 6. *For a random variable ξ, its cumulant-generating function is*

$$\Lambda_\xi(s) = \ln \mathbb{E} \exp(s\xi) \tag{19}$$

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

Definition 7. *The random variable ξ is* subgaussian *if there exists a constant C such that $\Lambda_\xi(t) \leq Ct^2$ 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 $\Lambda_\xi(t) \leq Ct^2$ on a neighborhood of 0). Note also that if $\mathbb{E}\xi = 0$ and $\|\xi\|_\infty < \infty$ then ξ is subgaussian.

Definition 8. *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*

$$f^*(y) = \sup_{x \in \mathbb{R}} \{xy - f(x)\} \tag{20}$$

Some basic properties of the Legendre transform include:

- If $f \leq g$ then $f^* \geq g^*$.
- 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).

For the definition of a large deviations principle (LDP), we refer to [37, Chapter 27].

Theorem 9. *Let ξ be a subgaussian random variable. For any integer $k \geq 1$ and any sequence m_n satisfying $\sqrt{n} \ll m_n \ll n$, the sequence*

$$\frac{1}{m_n}(\sigma_1(A_n), \ldots, \sigma_k(A_n)) \tag{21}$$

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

$$I(x) = \frac{|x|^2}{2} \inf_{s \in \mathbb{R}} \frac{\Lambda_\xi^*(s)}{s^2}. \tag{22}$$

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

$$\frac{1}{m_n}(\lambda_1(A_n), \ldots, \lambda_k(A_n)) \tag{23}$$

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

If A_n is the centered adjacency matrix of $\mathcal{G}(n, q)$ then it is covered by Theorem 9, where ξ is the random variable taking the values $-q$ and $1 - q$ with probabilities $1 - q$ and q respectively. In this case, we have

$$\Lambda_\xi^*(s) = D(q + s, q) := (q + s) \ln \frac{q + s}{q} + (1 - q - s) \ln \frac{1 - q - s}{1 - q}, \tag{24}$$

with the understanding that $\Lambda_\xi^*(s) = +\infty$ whenever $q + s \notin (0, 1)$. It is not hard to check – and we will do it in Section 5.5 – that $\frac{\Lambda_\xi^*(s)}{s^2}$ achieves its infimum at some $s \geq 0$ if and only if $q \leq \frac{1}{2}$.

In the case that $\frac{\Lambda_\xi^*(s)}{s^2}$ saturates its infimum only at negative s (corresponding to $q > \frac{1}{2}$ in the Bernoulli example), 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

$$\frac{1}{m_n^2} \ln \Pr\left(\sqrt{\sum_i \lambda_i^2(A_n)} > m_n t\right) \leq -\frac{t^2}{2} \inf_{s \in \mathbb{R}} \frac{\Lambda_\xi^*(s)}{s^2} + o(1) \tag{25}$$

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

$$\frac{1}{m_n^2} \ln \Pr\left(\sqrt{\sum_i \lambda_i^2(A_n)} > m_n t\right) \geq -\frac{t^2}{2} \inf_{s \geq 0} \frac{\Lambda_\xi^*(s)}{s^2} - o(1), \tag{26}$$

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

There are natural examples (including the Bernoulli example mentioned above) where $s^{-2}\Lambda_\xi^*(s)$ is increasing for $s \ge 0$. In this case,

$$\inf_{s\ge 0} s^{-2}\Lambda_\xi^*(s) = \lim_{s\to 0} s^{-2}\Lambda_\xi^*(s) = \frac{1}{2}(\Lambda_\xi^*)''(0) = \frac{1}{2\mathbb{E}\xi^2}, \tag{27}$$

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

$$\frac{1}{m_n^2}\ln \Pr\left(\lambda_1(A_n) > m_n t\right) \ge -\frac{t^2}{4\mathbb{E}\xi^2} - o(1). \tag{28}$$

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

Theorem 10. *In the setting of Theorem 9, if $\mathbb{E}\xi^3 < 0$ and $\lim_{s\to\infty} s^{-2}\Lambda_\xi(s) = 0$ then there exists some $\eta > 0$ such that for any $t > 0$,*

$$\lim_{n\to\infty} \frac{1}{m_n^2}\ln \Pr\left(\lambda_1(A_n) > m_n t\right) > -(1-\eta)\frac{t^2}{4\mathbb{E}\xi^2}. \tag{29}$$

In particular, the assumptions of Theorem 10 are satisfied for the (centered) Bernoulli random variable with $q > \frac{1}{2}$ mentioned above.

For our applications to random graphs, we require a version of Theorem 9 for random bits chosen without replacement. Specifically, we consider the Erdős-Rényi random graphs $\mathcal{G}(n,m)$, where m is an integer satisfying $|m - q\binom{n}{2}| = O(1)$ (and $q \in (0,1)$ is fixed).

Theorem 11. *Fix $q \in (0,1)$ and let A_n be the centered adjacency matrix of a $\mathcal{G}(n,m)$ random graph with $|m - q\binom{n}{2}| = O(1)$. For any integer $k \ge 1$ and any sequence m_n satisfying $\sqrt{n} \ll m_n \ll n$, the sequence*

$$\frac{1}{m_n}(\sigma_1(A_n), \ldots, \sigma_k(A_n)) \tag{30}$$

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

If, in addition, $q \le \frac{1}{2}$ then the sequence

$$\frac{1}{m_n}(\lambda_1(A_n), \ldots, \lambda_k(A_n)) \tag{31}$$

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

5. 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; a similar observation was also used in [31].

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 12. *For any symmetric matrix A,*

$$\left(\sum_{i=1}^{k} \max\{0, \lambda_i(A)\}^2\right)^{1/2} = \sup_{M \in \mathcal{M}_k^+} \langle A, M \rangle. \tag{32}$$

For any matrix A,

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

Proof. To prove the first claim, assume without loss of generality that $\lambda_1(A) > 0$ (if not, both sides are zero). Let $UDU^T = A$ be an eigen-decomposition of A (where D is diagonal and U is orthogonal), and assume without loss of generality that the diagonal elements of D are ordered as $\lambda_1(A) \geq \cdots \lambda_n(A)$. Let $\tilde{D}$ be the diagonal matrix with entries $\max\{0, \lambda_1(A)\}, \ldots, \max\{0, \lambda_k(A)\}, 0, \ldots, 0$, and define

$$M = \frac{U\tilde{D}U^T}{\|\tilde{D}\|_F} = \frac{U\tilde{D}U^T}{\left(\sum_{i=1}^k \max\{0, \lambda_i(A)\}^2\right)^{1/2}}. \tag{34}$$

Then $M \in \mathcal{M}_k^+$ and $\langle A, M \rangle = \|\tilde{D}\|_F = \left(\sum_{i=1}^k \max\{0, \lambda_i(A)\}^2\right)^{1/2}$. 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

$$\langle A, M \rangle \leq \langle A_+, M \rangle \leq \|A_+\|_F \|M\|_F \leq \|A_+\|_F = \sqrt{\sum_{i=1}^{k} \lambda_i(A_+)^2} = \sqrt{\sum_{i=1}^{k} \max\{0, \lambda_i(A)\}^2}. \tag{35}$$

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 9, it suffices to control

$$\Pr\left(\sup_{M \in \mathcal{M}_k^+} \langle A, M \rangle > tn^\alpha\right). \tag{36}$$

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

5.1. **The net argument.**

Definition 13. *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 $d(x, y) \leq \epsilon$.*

Lemma 14. *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,*

$$\sup_{M \in \mathcal{M}_k} \langle A, M \rangle \leq \frac{1}{1 - 2\epsilon} \sup_{N \in \mathcal{N}} \langle A, N \rangle. \tag{37}$$

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 hence we can write $N - M = \epsilon M_0 + \epsilon M_1$ for some $M_0, M_1 \in \mathcal{M}_k$. In other words, we can decompose

$$M = N + \epsilon M_0 + \epsilon M_1 \tag{38}$$

with $N \in \mathcal{N}$ and $M_0, M_1 \in \mathcal{M}_k$. It follows that

$$\langle A, N\rangle = \langle A, M\rangle - \epsilon\langle A, M_0\rangle - \epsilon\langle A, M_1\rangle \ge \langle A, M\rangle - 2\epsilon \sup_{M' \in \mathcal{M}_k} \langle A, M'\rangle,$$

and the claim follows. □

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. Such a bound can be found in [35, Lemma 3.1].

Lemma 15. *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 $(Ck/\epsilon)^{Cnk}$.*

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 16. *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 $t > 0$,*

$$\ln \Pr\left(\sum_{i=1}^k \sigma_i^2(A) > t\right) \le \sup_{M \in \mathcal{M}_k} \ln \Pr\left(\langle A, M\rangle \ge (1 - 2\epsilon)t\right) + O(nk \ln \frac{1}{\epsilon}). \tag{39}$$

Proof. For the first inequality, let $\mathcal{N}$ be an ϵ-net for $\mathcal{M}_k$ according to Lemma 15. By Lemma 12 and Lemma 14

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

By a union bound,

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

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

We remark that it is possible to prove a version of Proposition 16 for eigenvalues also, giving an upper bound on $\Pr(\sum \lambda_i^2(A) > t)$ in terms of

$$\sup_{M^+ \in \mathcal{M}_k^+} \Pr\left(\langle A, M^+\rangle \ge t\right). \tag{40}$$

This can in principle give 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.

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

$$\sup_{M \in \mathcal{M}_k} \ln \Pr\left(\langle A, M\rangle \geq t\right) \tag{41}$$

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

Lemma 17. *If ξ is subgaussian then*

$$4 \sup_{s\in\mathbb{R}} \frac{\Lambda_\xi(s)}{s^2} = \left(\inf_{s\in\mathbb{R}} \frac{\Lambda_\xi^*(u)}{u^2}\right)^{-1} < \infty. \tag{42}$$

Proof. The fact that $\sup_{s\in\mathbb{R}} \frac{\Lambda_\xi(s)}{s^2} < \infty$ 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, $\Lambda_\xi(s) \leq M_L(s)$ for all $s \in \mathbb{R}$. It follows that $\Lambda_\xi^*(u) \geq M_L^*(u) = \frac{u^2}{4L}$; in other words,

$$\frac{\Lambda_\xi^*(u)}{u^2} \geq \frac{M_L^*(u)}{u^2} = \frac{1}{4L} \tag{43}$$

for all u. This shows that

$$4 \sup_{s\in\mathbb{R}} \frac{\Lambda_\xi(s)}{s^2} \geq \left(\inf_{u\in\mathbb{R}} \frac{\Lambda_\xi^*(u)}{u^2}\right)^{-1}. \tag{44}$$

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

Proposition 18. *Let ξ be a random variable with everywhere-finite moment-generating function, and define*

$$\Lambda_\xi(s) = \ln \mathbb{E} \exp(s\xi) \tag{45}$$

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 $\ell^* = \sup_{s>0} \frac{\Lambda_\xi(s)}{s^2}$. Then

$$\sup_{\|M\|_F \leq 1} \Pr(\langle A, M\rangle > t) \leq \exp\left(-\frac{t^2}{8 \sup_{s>0} \frac{\Lambda_\xi(s)}{s^2}}\right) = \exp\left(-\frac{t^2}{2} \inf_{s>0} \frac{\Lambda_\xi^*(s)}{s^2}\right). \tag{46}$$

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 $\xi_1, \ldots, \xi_m$ 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 $s > 0$)

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

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 $\ell^* = \sup_{r>0} \frac{\Lambda_\xi(r)}{r^2}$ then

$$\sum_i \Lambda_\xi(sa_i) = \sum_i \frac{\Lambda_\xi(sa_i)}{(sa_i)^2}(sa_i)^2 \le s^2 \sum_i \ell^* a_i^2 \le \frac{s^2\ell^*}{2}. \tag{47}$$

Hence,

$$\Pr(\langle A, M\rangle > t) \le \exp\left(\frac{s^2\ell^*}{2} - \frac{st}{2}\right), \tag{48}$$

and the first claim follows by optimizing over s.

The second claim follows immediately from Lemma 17. $\square$

Putting Proposition 18 together with Proposition 16, we arrive at the following upper bound for singular values:

Corollary 19. *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*

$$\ln \Pr\left(\sqrt{\sum_{i=1}^k \sigma_i^2(A)} > t\right) \le -\frac{t^2L}{2} + O\left(nk \ln \frac{t^2L}{nk}\right). \tag{49}$$

Proof. We combine Proposition 18 and Proposition 16, setting $\epsilon = \frac{nk}{t^2L}$ (which is less than $\frac{1}{2}$ by assumption). This yields an upper bound of

$$-\frac{t^2L}{2} + O\left(nk + nk\ln\frac{t^2L}{nk}\right), \tag{50}$$

and the nk term can be absorbed in the final term. $\square$

Remark 20. *Note that the argument leading to Corollary 19 applies even when the entries ξ_{ij} are not identically distributed as long as $L \le \inf_s \frac{\Lambda_{ij}^*(s)}{s^2}$ for every i, j, where Λ_{ij} is the cumulant-generating function of ξ_{ij}.*

5.3. **Lower bound.** In this section, we give a lower bound that matches the upper bound of Corollary 19 whenever $\sqrt{n} \ll t \ll n$. The starting point is the lower bound of Cramér's theorem [37, Theorem 27.3]

Theorem 21. *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 for any $t > 0$,*

$$\frac{1}{m} \ln \Pr\left(\sum_{i=1}^m \xi_i > mt\right) \to -\Lambda^*(t) \tag{51}$$

as $m \to \infty$.

Proposition 22. *In the setting of Corollary 19, 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*

$$\ln \Pr\left(\sum_{i=1}^k w_i \sigma_i(A_n) > |w|\sqrt{t}\right) \geq -\frac{tL}{2} - o(t). \tag{52}$$

(Here, $|w|$ denotes $\sqrt{\sum_i w_i^2}$.) 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

$$\ln \Pr\left(\sum_{i=1}^k w_i \lambda_i(A_n) > |w|\sqrt{t}\right) \geq -\frac{tL}{2} - o(t). \tag{53}$$

Choosing arbitrary $w_1, \dots, w_k$ and applying the Cauchy-Schwarz inequality, Proposition 22 implies the same lower bounds on $\ln \Pr(\sum_i \sigma_i^2(A_n) > t)$ and $\ln \Pr(\sum_i \lambda_i^2(A_n) > t)$. In particular, it really is a lower bound that matches the upper bound of Corollary 19.

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 $w_i \leq \sigma_i(M) \leq w_i + 2|s_*|$. 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,

$$\Pr\left(\langle A, M\rangle > t\right) = \Pr\left(\operatorname{sgn}(s_*) \sum_{i=1}^m \xi_i > \frac{t}{2|s_*|}\right). \tag{54}$$

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

$$\frac{t}{s_*^2} = \frac{\sum_i w_i^2}{s_*^2} \leq \sum_i (\ell_i - 1)^2 = |\ell|^2 - 2\sum_i \ell_i + 2k. \tag{55}$$

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 von Neumann's trace inequality (see [36]) and the Cauchy-Schwarz inequality we have

$$\langle A, M\rangle \leq \sum_{i=1}^k \sigma_i(A)\sigma_i(M) \leq \sum_{i=1}^k \sigma_i(A)(w_i + 2s_*) \leq \sum_{i=1}^k \sigma_i(A)w_i + 2s_*\sqrt{k}\sqrt{\sum_{i=1}^k \sigma_i^2(A)}, \tag{56}$$

and hence

$$\Pr\left(\langle A, M\rangle > t\right) \leq \Pr\left(\sum_{i=1}^k \sigma_i(A)w_i > t - t^{2/3}\right) + \Pr\left(\sum_{i=1}^k \sigma_i^2(A) > \frac{t^{4/3}}{4s_*^2 k}\right). \tag{57}$$

By Corollary 19, the second probability is of order $\exp(-\Omega(t^{4/3}))$, and hence

$$\ln \Pr\left(\sum_{i=1}^k \sigma_i(A)w_i > t - t^{2/3}\right) \geq (1 - o(1)) \ln \Pr\left(\langle A, M\rangle > t\right) \geq -\frac{t\Lambda^*(s_*)}{2s_*^2} - o(t). \tag{58}$$

Substituting in $t = |w|\sqrt{t}$ in place of $t - t^{2/3}$, the extra error term can be absorbed in the $o(t)$ term.

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 (56) by

$$\langle A, M\rangle \leq \sum_{i=1}^k \lambda_i^+(A)\lambda_i(M) \leq \sum_{i=1}^k \lambda_i^+(A)(w_i + 2s_*) \leq \sum_{i=1}^k \lambda_i^+(A)w_i + 2s_*\sqrt{k}\sqrt{\sum_{i=1}^k \sigma_i^2(A)}, \tag{59}$$

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 22, namely that we have explicit candidates for extremal eigenvectors and singular vectors. We will state these just for the largest eigenvector, but of course they also hold in other situations.

Corollary 23. *Assume that $s \mapsto s^{-2}\Lambda^*(s)$ achieves its minimum at some $s_* \geq 0$. For $1 \ll t \ll n$, let $\ell = \lceil 1 + t/s_* \rceil$ and define $v \in \mathbb{R}^n$ by $v_1, \ldots, v_\ell = s_*^{1/2} t^{-1/2}$ and $v_{\ell+1}, \cdots, v_n = 0$. Then $|v| \leq 1 + o(1)$ and*

$$\ln \Pr(v^T A_n v \geq t) \geq -\frac{t^2 L}{2} - o(t^2). \tag{60}$$

Corollary 23 is immediate from the proof of Proposition 22, because in the case $k = 1$ and $w_1 = \sqrt{t}$, the M that we constructed in that proof is exactly $\sqrt{t}vv^T$. When we have extra quantitative control on the minimization of $\Lambda^*(s)/s$, it follows that the leading eigenvector must actually be close to the v described above. We show this in Section 8, restricted for simplicity to the Bernoulli setting.

5.4. **The LDP.** Putting together Corollary 19 and Proposition 22, we complete the proof of the LDP (Theorem 9). 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 $t = \inf_{x \in E} |x|$. If $t > 0$ then $\frac{1}{m_n}(\sigma_1(A_n), \dots, \sigma_k(A_n)) \in E$ implies that $\sum \sigma_i^2(A_n) > m_n^2 t^2$, and then Corollary 19 implies that

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

(And if $t = 0$ then the inequality above is trivially true.)

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 $\langle x, w\rangle \ge |w|^2$ and $|x|^2 \le |w|^2 + \epsilon$ then $x \in E$. Now, Proposition 22 implies that

$$\ln \Pr\left(\langle X, w\rangle \ge |w|^2\right) = \ln \Pr\left(\sum_i \sigma_i(A_n) w_i \ge m_n |w|^2\right) \ge -\frac{m_n^2 |w|^2 L}{2} - o(m_n^2) \tag{61}$$

On the other hand, Corollary 19 implies that

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

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

$$\ln \Pr(X \in E) \ge \ln \Pr\left(\langle X, w\rangle \ge |w|^2 \text{ and } |X|^2 \le |w|^2 + \epsilon\right) \ge -\frac{m_n^2 |w|^2 L}{2} - o(m_n^2). \tag{62}$$

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

The second part of Theorem 9 follows the exact same argument, only it uses the second part of Proposition 22.

5.5. **The case of $\mathcal{G}(n, m)$.** We next consider the case of Theorem 11. The first observation is that $q \le \frac{1}{2}$ if and only if $\Lambda^*(s)/s^2$ achieves its minimum at some non-negative s.

Lemma 24. *If $\xi = -q$ with probability $1 - q$ and $\xi = 1 - q$ with probability q then Λ^* (the convex conjugate of ξ's cumulant generating function) satisfies*

$$\inf_{s \in \mathbb{R}} \frac{\Lambda^*(s)}{s^2} = \frac{\ln \frac{1-q}{q}}{1 - 2q}, \tag{63}$$

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

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

$$D(r,q) = r \ln \frac{r}{q} + (1-r) \ln \frac{1-r}{1-q} \tag{64}$$

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

$$L(r) = \frac{D(r,q)}{(r-q)^2} = \frac{\Lambda^*(r-q)}{(r-q)^2} \tag{65}$$

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

$$L'(r) = -\frac{(q+r)\ln\frac{r}{q} + (2-q-r)\ln\frac{1-r}{1-q}}{(r-q)^3} =: -\frac{F(r)}{(r-q)^3}. \tag{66}$$

Then

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

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

$$x := q^{-2} \ln \frac{1}{1-q}, \qquad y := (1-q)^{-2} \ln \frac{1}{q}, \qquad \text{or } z := \frac{\ln\frac{1-q}{q}}{1-2q}. \tag{67}$$

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

$$q^2(1-2q)(z-x) = q^2 \ln \frac{1-q}{q} + (1-2q)\ln(1-q) = (1-q)^2 \ln(1-q) - q^2 \ln q. \tag{68}$$

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

Not only does Lemma 24 establish the unique minimizer, it shows that the behavior is locally quadratic around the minimizer. $D(q+s, q)$ is infinite outside the compact set $s \in [-q, 1-q]$, this also implies a quadratic lower bound on non-minimizers:

Corollary 25. *With the notation of Lemma 24, there is a constant $C = C(q)$ such that for every $s \in \mathbb{R}$,*

$$\frac{\Lambda^*(s)}{s^2} \geq \frac{\ln \frac{1-q}{q}}{1-2q} + C(s-(1-2q))^2.$$

To complete the proof of Theorem 11, it is enough to show that the upper bound of Corollary 19 and the lower bound of Proposition 22 still hold in this setting; then the proof of the LDP proceeds exactly as in the proof of Theorem 9. Checking Corollary 19 is trivial: recalling that A_n is the centered adjacency matrix of $\mathcal{G}(n,m)$ for $|m - q\binom{n}{2}| = O(1)$, we let $\tilde{A}_n$ be the centered adjacency matrix of $\mathcal{G}(n,q)$. 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,

$$\Pr(A_n \in F) = \Pr(\tilde{A}_n \in F \mid E) \leq \frac{\Pr(A_n \in F)}{\Pr(E)} \leq O(n \Pr(\tilde{A}_n \in F)). \tag{69}$$

In other words, $\ln \Pr(A_n \in F) \leq \ln \Pr(\tilde{A}_n \in F) + O(\ln n)$, and so Corollary 19 immediately implies the same upper bound for $\mathcal{G}(n,m)$.

For the lower bound, we need to look into the proof of Proposition 22. Recall that in the proof of Proposition 22, we constructed a matrix M with $O(t) = o(n^2)$ non-zero entries, all of which had the same value. For the $\mathcal{G}(n,q)$ 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 $H_{k,n,r}$ denotes a hypergeometric random variable with population size n, k successes, and r trials; and if $B_{q,r}$ denotes a binomial random variable with success probability q and r trials; then one easily shows using Stirling's approximation that

$$|\ln \Pr(H_{k,n,r} = s) - \ln \Pr(B_{k/n,r} = s)| = O(r^2/n). \tag{70}$$

In the setting of Proposition 22, the number of trials r is the number of non-zero elements in M, and since $r^2/n = O(t^2/n) = o(t)$, we have

$$\ln \Pr(\langle A_n, M\rangle > t) \geq \ln \Pr(\langle \tilde{A}_n, M\rangle > t) - o(t). \tag{71}$$

With this lower bound, we can follow the rest of the proof of Proposition 22 to complete the proof of Theorem 11.

6. Proof of Theorem 10

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 12, 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 22 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 $t = \sum_{i=1}^m t_i$,

$$\Pr\left(\sum_i a_i \xi_i > t\right) \ge \prod_{i=1}^m \Pr\left(\sum_{j=1}^{\lceil \tilde{m}_i \rceil} \xi_j > t_i/(\alpha b_i)\right) \tag{72}$$

and so Theorem 21 implies that if $\frac{t_i}{\alpha\beta m_i b_i} = \Theta(1)$ then

$$\ln \Pr\left(\sum_i a_i \xi_i > t\right) \ge -\beta \sum_i m_i \Lambda^*\left(\frac{t_i}{\alpha\beta m_i b_i}\right) - o\left(\beta \sum_i m_i\right). \tag{73}$$

Our goal will be to choose the parameters $m_i, b_i, \alpha, \beta,$ 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

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

With these choices, we have

$$\alpha^2 \beta = \frac{2}{\sum_i m_i b_i^2}, \tag{74}$$

meaning that

$$\sum_i a_i^2 = \alpha^2 \beta \sum_i m_i b_i^2 = 2 \tag{75}$$

and

$$\sum_i t_i = \alpha\beta \sum_i m_i b_i \Lambda'(b_i) = t. \tag{76}$$

(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 (73), we obtain

$$\ln \Pr\left(\sum_i a_i \xi_i > t\right) \ge -\frac{t^2}{2} \cdot \frac{\sum_i m_i b_i^2 \cdot \sum_i m_i \Lambda^*(\Lambda'(b_i))}{\left(\sum_i m_i b_i \Lambda'(b_i)\right)^2} - o(t^2), \tag{77}$$

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

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

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

$$\langle A, M\rangle = \sum_i a_i \xi_i, \tag{78}$$

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 (77), in terms of ϵ and δ. Before doing so, we observe some basic properties of the Legendre transform.

Lemma 26. *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

$$f^*(y) = \sup_z\{zy - f(z)\}, \tag{79}$$

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

$$f^*(f'(x)) = xf'(x) - f(x). \tag{80}$$

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 27. *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*

$$f^*(f'(\epsilon)) = f''(0)\frac{\epsilon^2}{2} + ((f^*)'''(0)(f'')^3(0) + 3f'''(0))\frac{\epsilon^3}{6} + O(\epsilon^4) \tag{81}$$

as $\epsilon \to 0$.

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

$$\frac{d}{d\epsilon}g(f'(\epsilon)) = g'(f'(\epsilon))f''(\epsilon), \tag{82}$$

and then

$$\frac{d^2}{d\epsilon^2}g(f'(\epsilon)) = g''(f'(\epsilon))(f''(\epsilon))^2 + g'(f'(\epsilon))f'''(\epsilon), \tag{83}$$

and finally

$$\frac{d^3}{d\epsilon^3}g(f'(\epsilon)) = g'''(f'(\epsilon))(f''(\epsilon))^3 + 3g''(f'(\epsilon))f''(\epsilon)f'''(\epsilon) + g'(f'(\epsilon))f''''(\epsilon). \tag{84}$$

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

$$g''(0)(f''(0))^2, \tag{85}$$

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

$$g'''(0)(f''(0))^3 + 3g''(0)f''(0)f'''(0). \tag{86}$$

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

Note that Λ satisfies the assumptions on f in Lemmas 26 (because we assumed that $\Lambda(s) = o(s^2)$) and 27 (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

$$L = \Lambda''(0) > 0. \tag{87}$$

Expanding out the parameters in (77), we have

$$\sum_i m_i b_i^2 = 1 + 2\frac{\epsilon}{\delta} + \frac{\epsilon^2}{\delta^2} \tag{88}$$

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 26 and our assumptions on Λ, the last term is $o(\epsilon^2)$. Applying Lemma 27 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

$$M = (\Lambda^*)'''(0)L^3 + 3\Lambda'''(0). \tag{89}$$

For the denominator in (77), 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 (77) 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

$$\ln \Pr(\lambda_1(A_n) > t) \geq \ln \Pr(\langle A, M\rangle > t\|M\|_F) \geq -(1-\eta)\frac{t^2}{4L} - o(t^2). \tag{90}$$

Replacing t by $m_n t$ and recalling that $L = \Lambda''(0) = \mathbb{E}\xi^2$ completes the proof of Theorem 10.

7. Back to cycle counts

We now turn to the proofs of Theorems 1 and 2. The proofs are very similar, so we devote most of this section to proving Theorem 1 and then indicate what changes must be made to obtain Theorem 2.

Our eigenvalue LDP (Theorem 9) allows us to control the cycle-count contribution from a constant number of very extreme eigenvalues, but in order to fully characterize the behavior of the cycle count, For this, we will use a deviation inequality from [33]:

Theorem 28. *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 $\delta \gg n^{-1}$,*

$$\Pr(|X_n - \mathbb{E}X_n| \geq \delta) \leq C \exp\left(-\frac{n^2\delta^2}{C\|\xi\|_\infty^2}\right). \tag{91}$$

Having controlled the bulk eigenvalues, we will use Corollary 19 to show that the cycle 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 11 is for the positive eigenvalues of centered matrices and we are interested in the negative eigenvalues here.

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

$$f_1(x) = \begin{cases} 0 & \text{if } x < 0 \\ x^k & \text{if } 0 \le x < \sqrt{K} \\ kK^{(k-1)/2}x - (k-1)K^{k/2} & \text{if } x \ge \sqrt{K} \end{cases} \tag{92}$$

and $f_2(x) = -f_1(-x)$. Then both f_1 and f_2 are $kK^{(k-1)/2}$-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 cycle 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 $K \ge C$ then*

$$\Pr\left(\frac{1}{n^k} \sum_{i:\lambda_i(A_n) \ge -\sqrt{Kn}} \lambda_i^k(A_n) < -t^k - Cn^{-1}\right) \le \exp\left(-\Omega\left(\frac{n^k t^{2k}}{K^{k-1}}\right)\right). \tag{93}$$

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

$$f_1(x) + f_2(x) \le \begin{cases} 0 & \text{if } x < -\sqrt{K} \\ x^k & \text{if } x \ge -\sqrt{K}. \end{cases} \tag{94}$$

Hence,

$$\sum_i (f_1 + f_2)(n^{-1/2}\lambda_i(A_n)) \le n^{-k/2} \sum_{i:\lambda_i(A_n) \ge -\sqrt{Kn}} \lambda_i^k(A_n). \tag{95}$$

Since $-f_2$ is convex, Theorem 28 applies to both f_1 and f_2, giving

$$\Pr\left(\frac{1}{n}\operatorname{tr}[(f_1+f_2)(n^{-1/2}A_n)] \le \frac{1}{n}\mathbb{E}\operatorname{tr}[(f_1+f_2)(n^{-1/2}A_n)] - s\right) \le 2\exp(-\Omega(n^2s^2/K^{k-1})) \tag{96}$$

whenever $s = \omega(K^{(k-1)/2}/n)$. Plugging in (96) gives

$$\Pr\left(\sum_{i:\lambda_i(A_n) \ge -\sqrt{Kn}} \lambda_i^k(A_n) \le n^{k/2}\mathbb{E}\operatorname{tr}(f_1+f_2)(n^{-1/2}A_n) - s\right) \le 2\exp\left(-\Omega\left(\frac{s^2}{K^{k-1}n^k}\right)\right). \tag{97}$$

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 $n^{-k/2}\mathbb{E}\operatorname{tr}(A_n^k)$. But note that

$$\begin{aligned}&|\operatorname{tr}[(f_1+f_2)(n^{-1/2}A_n)-n^{-k/2}A_n^k]|\\&\qquad\qquad\le n^{-k/2}\sum_{i:|\lambda_i|>\sqrt{Kn}}|\lambda_i(A_n)|^k\le n^{-k/2+1}|\sigma_1(A_n)|^k 1_{\{\sigma_1(A_n)|>\sqrt{Kn}\}},\end{aligned}$$

where $\sigma_1(A_n)$ is the largest singular value of A_n. Proposition 18 implies that if K is sufficiently large then $\mathbb{E}[|\sigma_1(A_n)|^k 1_{\{|\sigma_1(A_n)|>\sqrt{Kn}\}}]\le\exp(-\Omega(\sqrt{n}))$. Hence,

$$\Pr\left(\sum_{i:\lambda_i(A_n)\ge-\sqrt{Kn}}\lambda_i^k(A_n)\le\mathbb{E}\operatorname{tr}(A_n^k)-s-\exp(-\Omega(\sqrt{n}))\right)\le 2\exp\left(-\Omega\left(\frac{s^2}{K^{k-1}n^k}\right)\right).\tag{98}$$

Finally, note that $|\mathbb{E}\operatorname{tr}(A_n^k)|=O(n^{k-1})$ and set $s=t^kn^k$. □

We remark that the comparison between the exponents in Lemma 29 and the exponents in our eigenvalue LDP (Theorem 11) determines the range of deviations to which our cycle-count deviation bounds apply: we get sharp bounds whenever Lemma 29 ensures that the bulk contribution is smaller than the contribution of the most extreme eigenvalues. To that end, note that by Theorem 11, a single eigenvalue of order $-tn$ (which contributes t^k to the k-cycle density) carries an entropy cost of order t^2n^2; on the other hand, Lemma 29 shows that using the bulk eigenvalues to achieve the same t^k change in the k-cycle density costs $t^{2k}n^k$ in entropy. These costs cross over when $t^k\asymp n^{-\frac{k(2-k)}{2k-2}}$, but on the other hand applying Lemma 29 in this way also requires that $t^k\gg n^{-1}$. Therefore, we see that the bulk contribution is dominated by the extreme eigenvalue contribution whenever

$$t^k\gg n^{-\min\{1,\frac{k(2-k)}{2k-2})\}},$$

and the right hand side is $n^{-3/4}$ for $k=3$ and n^{-1} for $k\ge5$. This computation determines our critical exponent c_* given in (8).

7.2. **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 19 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 19, if $\sigma_i=\sigma_i(A)$ are the singular values of A then*

$$\ln\Pr\left(\sqrt{\sum_{\sigma_i>\sqrt{Kn}}\sigma_i^2}\ge t\right)\le-\frac{t^2L}{2}+O\left(\frac{t^2}{K}\ln K\right)\tag{99}$$

Proof. Set $k = \lceil t^2/(Kn) \rceil$ and observe that if $\sigma_1, \dots, \sigma_k \geq \sqrt{Kn}$ then $\sum_{i=1}^k \sigma_i^2 \geq t^2$. Hence, we either have

$$\sum_{\sigma_i > \sqrt{Kn}} \sigma_i^2 \leq \sum_{i=1}^k \sigma_i^2, \tag{100}$$

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

$$\ln \Pr\left(\sqrt{\sum_{\sigma_i > \sqrt{Kn}} \sigma_i^2} \geq t\right) \leq \ln \Pr\left(\sqrt{\sum_{i=1}^k \sigma_i^2} \geq t\right), \tag{101}$$

and we conclude by applying Corollary 19 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 19 to the $\mathcal{G}(n,m)$ case, namely that a $\mathcal{G}(n,m)$ graph can be obtained by conditioning a $\mathcal{G}(n,q)$ graph on an event of $\Omega(n^{-1})$ probability. □

Corollary 30 for extends to the case of a $\mathcal{G}(n,m)$ graph by the same argument that was used to extend Corollary 19 in the proof of Theorem 11. Namely, a $\mathcal{G}(n,m)$ graph can be obtained by conditioning a $\mathcal{G}(n,q)$ graph on an event of $\Omega(n^{-1})$ probability, and the extra factor n introduced by the conditioning is of smaller order. Applying Corollary 30 to a centered $\mathcal{G}(n,q)$ adjacency matrix for $q = m/\binom{n}{2}$, and then applying Lemma 24 to get the explicit expression for L, we obtain the following bound:

Corollary 31. *If A is the centered adjacency matrix of a $\mathcal{G}(n,m)$ random graph, let $p = m/\binom{n}{2}$ and let*

$$L = \frac{\ln \frac{1-p}{p}}{1-2p}.$$

Let $\sigma_i = \sigma_i(A)$ be the singular values of A. Then for any fixed K, if $t \gg \sqrt{n}$

$$\ln \Pr\left(\sqrt{\sum_{\sigma_i > \sqrt{Kn}} \sigma_i^2} \geq t\right) \leq -\frac{t^2 L}{2} + O\left(\frac{t^2}{K} \ln K\right) \tag{102}$$

7.3. **The upper bound in Theorem 1.** Let A be the adjacency matrix of a $\mathcal{G}(n,m)$ graph and recall that $\tau_k(A) = \frac{\mathrm{tr}[A^k]}{n^k} + O(1/n)$. Let $\tilde{A} = A - \mathbb{E}A$; by Corollary 5,

$$\begin{aligned} \Pr(\tau_k(A) \leq p^k - t^k) &= \Pr(\mathrm{tr}[A^k] \leq n^k p^k - n^k t^k + O(n^{k-1})) \\ &\leq \Pr(\mathrm{tr}[\tilde{A}^k] \leq -n^k t^k + O(n^{k-1})) + \Pr(\|\tilde{A}\|_{\mathrm{op}} \geq \Omega(n)). \end{aligned} \tag{103}$$

Writing out $\mathrm{tr}[\tilde{A}^k] = \sum_i \lambda_i^k(\tilde{A})$, choose $K = \omega(1)$ and $\epsilon = o(1)$ such that $K^{k-1}/\epsilon^{2/k} = o(n^{k-2}t^{2k-2})$; this is possible because $t^k \gg n^{-c_*}$ implies that $n^{k-2}t^{2k-2} \gg 1$. Applying Lemma 29 to $\tilde{A}$ gives

$$\Pr\left(n^{-k} \sum_{i:\lambda_i \geq -\sqrt{Kn}} \lambda_i^k(\tilde{A}) < -\epsilon t^k\right) \leq \exp\left(-\Omega\left(\frac{\epsilon^{2/k} t^{2k} n^k}{K^{k-1}}\right)\right) = \exp(-\omega(n^2 t^2)). \tag{104}$$

On the other hand, Jensen's inequality implies that

$$\left|\sum_{i:\lambda_i<-\sqrt{Kn}} \lambda_i^k\right| \le \left(\sum_{i:\lambda_i<-\sqrt{Kn}} \lambda_i^2\right)^{k/2} \le \left(\sum_{i:\sigma_i>\sqrt{Kn}} \sigma_i^2\right)^{k/2}, \tag{105}$$

where $\lambda_i = \lambda_i(\tilde{A})$ and $\sigma_i = \sigma_i(\tilde{A})$. Recall here that $L = \inf_{s\in\mathbb{R}} \frac{\Lambda^*(s)}{s^2}$, where Λ is the cumulant-generating function of a centered Bernoulli random variable with success probability p. Lemma 24 (with $q = 1-p$) implies that $L = \frac{\ln\frac{p}{1-p}}{2p-1}$. By Corollary 31 (and taking into account the fact that $\epsilon = o(1)$ and $K = \omega(1)$),

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

Combined with (104), this yields

$$\ln\Pr\left(\mathrm{tr}[\tilde{A}^k] \le -t^k n^k\right) \le -\frac{Lt^2n^2}{2}(1+o(1)). \tag{106}$$

Now we apply (103), noting that $n^k t^k = \omega(n^{k-1})$, and so $n^k t^k + O(n^{k-1}) = n^k t^k(1+o(1))$, to get

$$\ln\Pr(\tau_k(A) \le p^k - t^k) \le \max\left\{-\frac{Lt^2n^2}{2}(1+o(1)), \ln\Pr(\|\tilde{A}\|_{\mathrm{op}} \ge \Omega(n))\right\}. \tag{107}$$

By Theorem 11, the second term in the maximum is of order $-\Theta(n^2)$ and so the first term wins.

This completes the proof of the upper bound in Theorem 1 but let us also note two other facts that we can easily extract from the proof. From (104) we see that only the extremely negative eigenvalues contribute to the cycle deviation:

Corollary 32. *Conditioned on* $\tau_k(A) \le p^k - t^k$, $\sum_{i:\lambda_i\le-\Omega(\sqrt{n})} \lambda_i^k(\tilde{A}) \le -t^k n^k(1-o(1))$ *with high probability.*

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

Corollary 33. *Conditioned on* $\tau_k(A) \le p^k - t^k$, *if* $d_1, \dots, d_n$ *are the vertex degrees of the graph then with high probability*

$$\sum_i (d_i - pn)^2 = o(t^k n^3). \tag{108}$$

Proof. In the proof of the upper bound of Theorem 1, recall that $\ln\Pr(\|\tilde{A}\|_{\mathrm{op}} \ge \Omega(n)) \ll \ln\Pr(\tau_k(A) \le p^k - t^k)$, and it follows that conditioned on $\tau_k(A) \le p^k - t^k$ we have $\|\tilde{A}\|_{\mathrm{op}} = o(n)$ with high probability. Since $\Pr(\tau_k(A) \le p^k - t^k(1+\epsilon)) \ll \Pr(\tau_k(A) \le p^k - t^k)$, we also have $\tau_k(A) \geq p^k - t^k(1+\epsilon)$ with high conditional probability. But on the event that $\|\tilde{A}\|_{\mathrm{op}} \leq \epsilon n$ and $\tau_k(A) \geq p^k - t^k(1+\epsilon)$, Lemma 4 implies that

$$n^{k-3} \sum_i (d_i - pn)^2 \leq \epsilon t^k n^k,$$

and the claim follows. □

7.4. **The lower bound in Theorem 1.** The idea here is to partition the adjacency matrix into blocks, and then consider the event that certain prescribed numbers of edges are present in each block. By choosing all parameters correctly, we can ensure that this event has the correct probability, and also that on this event the cycle density will behave as desired.

Recall that $L = \sup_s \frac{\Lambda^*(s)}{s^2}$ and that $s_* = 2p - 1$ is the maximizing value of s. Let ℓ be the closest integer to tn/s_* and let $\xi_1, \dots, \xi_{\binom{n}{2}}$ be some ordering of the upper diagonal of $\tilde{A}$. Let U_{11} be the collection of i for which ξ_i is in the upper-left $\ell \times \ell$ submatrix; let U_{12} be the collection of i for which ξ_i is in the upper-right $\ell \times (n-\ell)$ submatrix; and let U_{22} be the remaining indices. Define z by $\ell = zn$, and note that $z = (1 + O(1/n))t/s_*$. Now let $S_* = \lfloor s_* |U_{11}| \rfloor$ and $T_* = \lceil z |U_{12}| s_* \rceil$, and let Ω be the event that

$$\frac{1}{|U_{11}|} \sum_{i \in U_{11}} \xi_i = S_* \tag{109}$$

$$\frac{1}{|U_{12}|} \sum_{i \in U_{12}} \xi_i = T_* \tag{110}$$

We claim that $\ln \Pr(\Omega) \geq -\frac{t^2 n^2 L}{2}(1 + o(1))$, and that conditioned on Ω, $\tau_k(G) \leq p^k - t^k$ with non-negligible probability. Together, these imply the lower bound of Theorem 1.

Lemma 34.

$$\ln \Pr(\Omega) \geq -\frac{t^2 n^2 L}{2}(1 + o(1)).$$

Proof. Let Ω_1 be the event of (109) and let Ω_2 be the event of (110). These events can be described simply in terms of hypergeometric random variables. Indeed, $\sum_{i \in U_{11}} (\xi_i + p)$ is a hypergeometric random variable with $\binom{\ell}{2}$ trials, and a population of size $\binom{n}{2}$ containing m successes; therefore Ω_1 is just the event that this hypergeometric variable takes a particular value. Conditioned on Ω_1, $\sum_{i \in U_{12}} (\xi_i + p)$ is a hypergeometric random variable with $\ell(n-\ell)$ trials, and a population of size $\binom{n}{2} - \binom{\ell}{2}$ containing $m - S_*$ successes; the event Ω_2 is just the event that this hypergeometric variable takes a particular value.

These hypergeometric probabilities can be computed explicitly; we will make use of the approximation that comes simply from applying Stirling's approximation to the explicit computation (see, e.g., [38, Lemma 2.1.33]): if Z is a hypergeometric random variable with r trials from a population of size R with αR successes, then for any integer b in the range of Z,

$$\frac{1}{r} \ln \Pr(Z = b) = -D(b/r, \alpha) - \frac{1 - r/R}{r/R} D\left(\frac{\alpha - b/R}{1 - r/R}, \alpha\right) + O\left(\frac{\ln R}{r}\right) \tag{111}$$

where $D(q+s,q) = (q+s)\ln\frac{q+s}{q} + (1-q-s)\ln\frac{1-q-s}{1-q}$ is, as before, the Legendre transform of a centered Bernoulli variable's cumulant generating function.

Applying (111) to Ω_1, since $\ln n \ll \ell \ll n$ and since $D(\alpha+\epsilon,\alpha) = \Theta(\epsilon^2)$, we have

$$\frac{1}{|U_{11}|}\Pr(\Omega_1) \to -D(p+s^*,p) = -L,$$

and hence

$$\ln\Pr(\Omega_1) = -(1+o(1))\frac{t^2n^2L}{2}.$$

Since $\Omega = \Omega_1 \cap \Omega_2$, it suffices to show that

$$\Pr(\Omega_2 \mid \Omega_1) = \exp(-o(t^2n^2)).$$

Recall that conditioned on Ω_1, $\sum_{i\in U_{12}}(\xi_i + p)$ is hypergeometric with $\Theta(zn^2)$ trials, a population size of $\Theta(n^2)$, and $m - O(z^2)$ successes. The event Ω_2 is asking for this hypergeometric variable to deviate from its mean (which is of order $\Theta(zn^2)$) by a fixed quantity of smaller order, namely $\Theta(z^2n^2)$. By (111),

$$\frac{1}{|U_{12}|}\ln\Pr(\Omega_2 \mid \Omega_1) = -D(p+\Theta(z),p) + o(z) = -o(z).$$

Therefore, $\Pr(\Omega_2 \mid \Omega_1) = \exp(-o(z^2n^2)) = \exp(-o(t^2n^2))$. □

Next, we show that conditioned on Ω, G has fewer cycles. For ease of notation, let us first describe the conditional distribution of G given Ω in terms of different parameters. Let $\ell = zn$ for $n^{-2/3} \ll z \ll 1$, and fix $0 < q < p$. Consider a random graph G with $m = p\binom{n}{2}$ edges, and let A be its adjacency matrix. Suppose that $(p-q)\binom{\ell}{2}$ of these edges are uniformly distributed on the upper diagonal of the top-left $\ell\times\ell$ block of A, $(p+\frac{z}{1-z}q)\ell(n-\ell)+O(1)$ are uniformly distributed on the top-right $\ell\times(n-\ell)$ block, and $(p-\frac{z^2}{(1-z)^2}q)\binom{n-\ell}{2}+O(1)$ are uniformly distributed on the remaining part of the upper-diagonal. The $O(1)$ error terms ensure that it is possible to satisfy the constraints with integer numbers of edges, and these error terms are also compatible with the requirement that there are $p\binom{n}{2}$ edges in total. Finally, note that the distribution of G conditioned on Ω is the same as the distribution we have described above, for some q within $\Theta(1/\ell^2)$ of s_*.

Lemma 35. *For the random graph G described above,*

$$\mathbb{E}\tau_k(G) = p^k - z^kq^k + o(z^k)$$

and

$$\mathrm{Var}(\tau_k(G)) = O(n^{-2}).$$

In particular, if $z^kq^k = \omega(n^{-1})$ then $\tau_k(G) \le p^k - z^kq^k + o(z^k)$ w.h.p.

Recalling that $z^kq^k \ge (1+o(1))t^kn^k$, Lemma 35 completes the proof of the lower bound of Theorem 1, after replacing t by $(1-o(1))t$.

Proof. To compute the expected number of cycles, let B the the $n\times n$ block matrix that agrees with $\mathbb{E}A$ except on the diagonal. That is, B takes the value $p-q$ on the top-left $\ell\times\ell$ block, the value $(p+\frac{z}{1-z}q)+O(n^{-2}z^{-1})$ on the top-right $\ell\times(n-\ell)$ block, and the value $(p-\frac{z^2}{(1-z)^2}q)+O(n^{-2})$ on the bottom $(n-\ell)\times(n-\ell)$ block. Then B has rank-2, and it is well-approximated by the rank-2 matrix $p\mathbf{1} - qww^T$, where w has ones in the first zn entries, and takes the value $-z/(1-z)$ in the other entries. More precisely,

$$\|B - (p\mathbf{1} - qww^T)\|_F = O(n^{-1}z^{-1/2}) = o(1),$$

with the main contribution coming from the $\Theta(zn^2)$ entries of size $O(z^{-1}n^{-2})$. Since $\mathbf{1}$ and w are orthogonal, $p\mathbf{1} - qww^T$ has eigenvalues pn and $-q|w|^2 = -qzn + O(z^2 n)$. By Weyl's eigenvalue inequalities, B has eigenvalues $pn + o(1)$ and $-qzn + o(1)$. Therefore $\operatorname{tr} B^k = p^k n^k - q^k z^k n^k + O(n^{k-1})$.

Next, consider $\operatorname{tr}[(\mathbb{E}A)^k]$. Recalling that $\mathbb{E}A$ agrees with B except on the diagonal (because $\mathbb{E}A$ is zero on the diagonal and B is not), we have $\|\mathbb{E}A - B\|_{\mathrm{op}} = O(1)$, and so Weyl's eigenvalue inequalities imply that $\mathbb{E}A$ has an eigenvalue of $pn + O(1)$, an eigenvalue of $-qzn + O(1)$, and its remaining eigenvalues are bounded. Therefore, $\operatorname{tr}[(\mathbb{E}A)^k] = p^k n^k - q^k z^k n^k + O(n^{k-1})$.

To compare $\operatorname{tr}[(\mathbb{E}A)^k]$ to $\mathbb{E}\tau_k(G)$, expand $\operatorname{tr}[(\mathbb{E}A)^k]$ in terms of closed walks of length k: let Γ_k be the set of $(k+1)$-tuples $v_1, \dots, v_{k+1}$ with $v_{k+1} = v_1$ and $v_i \neq v_{i+1}$ for all i. Then

$$\operatorname{tr}[(\mathbb{E}A)^k] = \sum_{(v_1,\dots,v_{k+1}) \in \Gamma_k} \prod_{i=1}^k (\mathbb{E}A)_{v_i, v_i+1}. \tag{112}$$

Let $\tilde{\Gamma}_k$ be the set of $(k+1)$-tuples $v_1, \dots v_{k+1}$ in Γ_k such that $v_1, \dots, v_k$ are distinct. Then $|\Gamma_k - \tilde{\Gamma}_k| = O(n^{k-1})$. Since each summand in (112) is bounded,

$$\operatorname{tr}[(\mathbb{E}A)^k] = \sum_{(v_1,\dots,v_{k+1}) \in \tilde{\Gamma}_k} \prod_{i=1}^k (\mathbb{E}A)_{v_i, v_i+1} + O(n^{k-1}).$$

On the other hand,

$$\binom{n}{k} \mathbb{E}\tau_k(G) = \sum_{(v_1,\dots,v_{k+1}) \in \tilde{\Gamma}_k} \Pr(\{v_i, v_{i+1}\} \in E(G) \text{ for all } i).$$

For each i, $\Pr(\{v_i, v_{i+1}\} \in E(G)) = (\mathbb{E}A)_{v_i, v_{i+1}}$. Because the edges are chosen without replacement these terms are not independent. However, we always have the inequality

$$\Pr(\{v_i, v_i + 1\} \in E(G) \mid \{v_1, v_2\} \in E(G), \dots, \{v_{i-1}, v_i\} \in E(G) \leq (\mathbb{E}A)_{v_i, v_{i+1}}.$$

Therefore,

$$\binom{n}{k} \mathbb{E}\tau_k(G) \leq \sum_{(v_1,\dots,v_{k+1}) \in \tilde{\Gamma}_k} \prod_{i=1}^k (\mathbb{E}A)_{v_i, v_i+1} = \operatorname{tr}[(\mathbb{E}A)^k] + O(n^{k-1}) = p^k n^k - q^k z^k n^k + O(n^{k-1}).$$

This proves the claim about the expectation.

Next, we consider the variance of the cycle density. For an ordered k-tuple $S \subset V(G)$, let T_S be the event that the vertices in S form a k-cycle. Note that because the edges are drawn without replacement, if S_1 and S_2 do not share an edge then T_{S_1} and T_{S_2} are non-positively correlated. Therefore,

$$\operatorname{Var}(T(G)) = \sum_{S_1, S_2} \operatorname{Cov}(T_{S_1}, T_{S_2}) = \sum_{S_1, S_2 : |S_1 \cap S_2| \geq 2} \operatorname{Cov}(T_{S_1}, T_{S_2}).$$

There are at most n^{2k-2} elements in the sum, and each is bounded by 1. Therefore, $\mathrm{Var}(T(G)) \le n^{2k-2}$ and so $\mathrm{Var}(\tau(G)) = O(n^{-2})$. □

7.5. **The two extreme eigenvalues.** In proving the upper bound on $\Pr(\tau_k(A) \le p^k - t^k)$, we applied the inequality $\sum_i |a_i|^k \le (\sum_i a_i^2)^{k/2}$ 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. In what follows, we write $\|a\|_p^p$ for $\sum_i |a_i|^p$.

Lemma 36. *Let $a_1, \dots$ be a sequence of non-negative numbers, in non-increasing order. For $\epsilon > 0$ and $k \ge 3$, if*

$$\sum_{i \ge 2} a_i^k \ge \epsilon a_1^k \tag{113}$$

then

$$\|a\|_2^2 \ge (1+\epsilon)^{1/k} \|a\|_k^2. \tag{114}$$

Proof. If $\sum_{i\ge 2} a_i^k \ge \epsilon a_1^k$ then $\|a\|_\infty^k = a_1^k \le \frac{\|a\|_k^k}{1+\epsilon}$. Then $\|a\|_k^k \le \|a\|_\infty^{k-2}\|a\|_2^2 \le (1+\epsilon)^{-(k-2)/k}\|a\|_k^{k-2}\|a\|_2^2$, and the claim follows. □

Applying Lemma 36 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 37. *In the setting of Theorem 1, for any $\epsilon > 0$, conditioned on $\tau_k(A) \le p^k - t^k$ we have*

$$\lambda_n^k(\tilde{A}) \le -(1-\epsilon)t^k n^k \text{ and } \lambda_{n-1}^k(\tilde{A}) \ge -\epsilon t^k n^k \tag{115}$$

with high probability.

Proof. Let $S = \{i : \lambda_i(\tilde{A}) \le -\Omega(\sqrt{n})\}$. By Corollary 32, for any $\delta > 0$, conditioned on $\tau_k(A) \le p^k - t^k$ we have

$$\sum_{i \in S} \lambda_i^k(\tilde{A}) \le -(1-\delta)t^k n^k \tag{116}$$

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

$$\Pr\left(\sum_{i\in S} \lambda_i^k(\tilde{A}) \le -(1-\delta)t^k n^k \text{ and } \lambda_n^k(\tilde{A}) > -(1-\delta-\epsilon)t^k n^k \text{ and } \sum_{i\in S\setminus\{n\}} \lambda_i^k(\tilde{A}) \le -\epsilon t^k n^k\right) \tag{117}$$

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

Indeed, applying Lemma 36 to the sequence of $|\lambda_i|$ for $i \in S$, we see that if

$$\sum_{i\in S} \lambda_i^k(\tilde{A}) \le -(1-\delta)t^k n^k \text{ and } \lambda_n^k(\tilde{A}) > -(1-\delta-\epsilon)t^k n^k \text{ and } \sum_{i\in S\setminus\{n\}} \lambda_i^k(\tilde{A}) \le -\epsilon t^k n^k \tag{118}$$

then

$$\sum_{i\in S} \lambda_i^2(\tilde{A}) \geq (1+\epsilon)^{1/k}(1-\delta)t^2n^2 \geq (1+\Omega(\epsilon))t^2n^2, \tag{119}$$

where the last inequality follows by choosing a sufficiently small $\delta = \Omega(\epsilon)$. But Corollary 31 implies that

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

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

Note that although we have been focussing on the smallest (i.e. negative, with large magnitude) eigenvalues, this same argument tells us about the largest eigenvalues also: if $\lambda_1(\tilde{A}) \geq \epsilon^{1/k}tn$, then in order to have $\sum_i \lambda_i^k(\tilde{A}) \leq -(1-o(1))t^kn^k$ we would need $\sum_{i=2}^n \lambda_i^k(\tilde{A}) \leq -(1-\epsilon-o(1))t^kn^k$, which by the argument above has probability $\exp(-(1+\Omega(\epsilon))Lt^2n^2/2) = o(\Pr(\tau_k(A) \leq p^k - t^k))$. Therefore, we obtain the following bound on the largest eigenvalue:

Corollary 38. *In the setting of Theorem 1, conditioned on $\tau_k(A) \leq p^k - t^k$, with high probability $\lambda_1(\tilde{A}) = o(tn)$.*

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

$$\lambda_{n-1}(A) \geq \lambda_{n-1}(\tilde{A}) - p, \tag{120}$$

and so $\lambda_{n-1}(\tilde{A}) \geq -o(tn)$ implies the same for $\lambda_{n-1}(A)$. For λ_n, let v be a unit eigenvector of $\tilde{A}$ with eigenvalue $\lambda_n(\tilde{A})$. By Corollary 33, with high (conditional on $\tau_k(A) \leq p^k - t^k$) probability, $|\tilde{A}1|^2 = o(t^kn^3)$, where 1 denotes the all-ones vector. On this event, expand 1 in the basis of eigenvectors of $\tilde{A}$ to see that $|\tilde{A}1|^2 \geq \langle 1, v\rangle^2 \lambda_n(\tilde{A})^2$. Therefore $\langle 1, v\rangle^2 \leq o(t^kn^3\lambda_n(\tilde{A})^{-2}) = o(t^{k-2}n) = o(tn)$. Now, $\langle A, vv^T\rangle \leq \langle \tilde{A}, vv^T\rangle + p\langle 1, v\rangle^2 = \lambda_n(\tilde{A}) + o(tn)$ and by considering v as a potential eigenvector of A, it follows that $\lambda_n(A) \leq \lambda_n(\tilde{A}) + o(tn)$. This completes the proof of Theorem 1.

7.6. **Theorem 2.** Like Theorem 1, Theorem 2 has three elements: an upper bound on the probability of a moderate deviation, a lower bound, and a bound on the most negative eigenvalue of the adjacency matrix.

The upper bound is proved exactly as in the proof of Theorem 1. The singular values of the eigenvalues are controlled by the rate function involving $\inf_{s\in\mathbb{R}} \frac{\Lambda^*(s)}{s^2}$, which we have already established to be $\frac{\ln\frac{1-p}{p}}{2(1-2p)}$. Upper bounds on singular values then give upper bounds on eigenvalues. The entire argument is independent of whether $p \geq \frac{1}{2}$ or $p < \frac{1}{2}$.

The proof of the lower bound in Theorem 2 is similar to that of the lower bound in Theorem 1, except that we use the vector $v = (\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}, -\frac{1}{\sqrt{n}}, \ldots, -\frac{1}{\sqrt{n}})$. For this v, Cramér's theorem shows that $\ln\Pr(\langle \tilde{A}, vv^T\rangle \leq -tn) \geq -\frac{t^2n^2}{2p(1-p)}(1+o(1))$, and the rest of the proof proceeds as before.

For the claim about the eigenvalue, we use Lemma 36: fix $\eta > 0$ and $K > 0$ and consider the event Ω on which

$$\sum_{\lambda_i(\tilde{A}) \leq -\sqrt{Kn}} \lambda_i^k(\tilde{A}) \leq -t^k n^k$$

but $\lambda_n^k(\tilde{A}) \geq -\frac{1}{1+\eta} t^k n^k$. According to Lemma 36, on this event we have

$$\sum_{\lambda_i(\tilde{A}) \leq -\sqrt{Kn}} \lambda_i^2(\tilde{A}) \geq (1+\eta)^{1/k} t^2 n^2.$$

By Corollary 31 (for a sufficiently slowly growing $K = \omega(1)$),

$$\ln \Pr(\Omega) \leq -\frac{t^2 n^2 (1+\eta)^{1/k} \ln \frac{p}{1-p}}{2(2p-1)} + o(t^2 n^2),$$

which, for sufficiently large η (depending on p) implies that

$$\ln \Pr(\Omega) \leq -(1+\Omega(1)) \frac{t^2 n^2}{2p(1-p)}.$$

It follows from the lower bound in Theorem 2 that $\Pr(\Omega \mid \tau_k \leq p^k - t^k) \to 0$. Together with Lemma 29 – which shows that eigenvalues larger than $-\sqrt{Kn}$ are unlikely to contribute – this implies that $\lambda_n^k \leq -\frac{1}{1+\eta} t^k n^k$ with high probability given $\tau_k \leq p^k - t^k$. (The main difference here compared to the proof of Theorem 1 is that because we lack matching upper and lower bounds on the log-probabilities, we cannot take $\eta \approx 0$.)

8. The conditional structure

In our upper bounds on eigenvalue deviation probabilities, we identified a candidate worst-case eigenvector: specifically, one that takes a certain non-zero value on $\Theta(tn)$ coordinates and zero elsewhere. In order to identify the conditional structure of this graph, we need to show that this candidate eigenvector is essentially the only one: every candidate eigenvector that has a comparable deviation probability is close to the one we identified.

The first step is to characterize the values that give the worst-case result in our Hoeffding-type bounds. For the rest of this section, fix p and let $\Lambda(u) = \ln(pe^{u(1-p)} + (1-p)e^{-up})$ be the cumulant-generating function of a centered, q-biased Bernoulli variable.

Lemma 39. *The function $\frac{\Lambda(u)}{u^2}$ has a unique maximizer u_*, and there is a constant $c = c(p) > 0$ such that for every $u \in \mathbb{R}$,*

$$\frac{\Lambda(u)}{u^2} \leq \frac{\Lambda(u_*)}{(u_*)^2} - c \min\{1, (u-u_*)^2\}.$$

Proof. Let $F(u) = \frac{\Lambda(u)}{u^2}$ (which is continuously defined and differentiable at zero by taking limits). Note that $\Lambda(u)$ is asymptotic to $u(1-p)$ as $u \to \infty$ and asymptotic to $-up$ as $u \to -\infty$. In particular, $F(u) \to 0$ as $u \to \pm\infty$, and since F is continuous on $\mathbb{R}$ with $F(0) = 0$ we see that it achieves a maximum at (possibly more than one) $u \in R$. Let $\ell^* = \sup_u \frac{\Lambda(u)}{u^2}$, and suppose that u_* achieves the maximum. Then Λ and $u \mapsto \ell^* u^2$ have the same tangent at u_*. Since Λ is convex, the function $x \mapsto 2u_* x - \ell^*(u_*)^2$ touches Λ from below at u_*, and it follows that $\Lambda^*(2\ell^* u_*) = \ell^*(u_*)^2$. Or in other words, $\Lambda^*(s_*) = \frac{(s_*)^2}{4\ell^*}$ for

$s_* = 2\ell^* u_*$. Now recall from Lemma 17 that $\frac{1}{4\ell^*} = \inf_y \frac{\Lambda^*(y)}{y^2}$. It follows that for every u_* at which $\Lambda(u)/u^2$ achieves its maximum, there is a s_* at which $\Lambda^*(y)/y^2$ achieves its minimum. By Lemma 24, $\Lambda^*(y)/y^2$ has a unique minimizer and it follows that $\Lambda(u)/u^2$ has a unique maximizer.

To see that $\Lambda(u)/u^2$ is locally quadratic near u_*, note that $\Lambda''(u_*)(\Lambda^*)''(s_*) = 1$. By Corollary 25, $(\Lambda^*)''(s_*) > \frac{1}{2\ell^*}$ and it follows that $\Lambda''(u_*) < 2\ell^*$ and so $F(u)$ is locally quadratic near u_*. And since $F > 0$ at its unique maximizer and $F(u) \to 0$ at $\pm\infty$, it follows that $F(u) \le F(u_*) - c\min\{1, (u - u_*)^2\}$ for some $c > 0$ and all $u \in \mathbb{R}$. □

With this extra information on the maximizer of $\Lambda(u)/u^2$, we revisit the Hoeffding-type argument of Proposition 18: in order for a matrix M to get close to the upper bound of Proposition 18, most of the contribution to $\|M\|_F$ must come from entries that are close to the "ideal value". Recall that $s_* = 2p - 1$ is the unique minimizer of $\frac{\Lambda^*(s)}{s^2}$, where Λ^* is the convex conjugate of Λ. The matrix M that we constructed in Proposition 18 had all of its entries being either zero or s_*/t; the next result shows that this is essentially necessary.

Proposition 40. *With ξ the centered, p-biased Bernoulli variable as above, let A be the symmetric random matrix with zero diagonal, and with upper-diagonal elements distributed independently according to ξ. For any $\|M\|_F \le 1$ and $t > 0$,*

$$\Pr(\langle A, M\rangle \ge t) \le \exp\left(-\frac{t^2 L}{2} - \Omega\Big(t^2 \sum_i \min\{a_i^2, (a_i - s_*/t)^2\}\Big)\right),$$

where a_i are the upper-diagonal entries of M.

Proof. As in Proposition 18, for any $s \in \mathbb{R}$ we have

$$\Pr(\langle A, M\rangle \ge t) \le \exp\left(\sum_i \Lambda(sa_i) - st/2\right).$$

By Lemma 39,

$$\sum_i \Lambda(sa_i) = \sum_i \frac{\Lambda(sa_i)}{(sa_i)^2}(sa_i)^2 \le s^2 \ell^* \sum_i a_i^2 - cs^2 \sum_i a_i^2 \min\{1, (a_i s - u_*)^2\},$$

where $\ell^* = \sup_u \frac{\Lambda(u)}{u^2}$ and $u_* = \frac{s_*}{2\ell^*} = 2s_* L$ is the unique maximizer. (Recalling that s_* is the unique minimizer of $\frac{\Lambda^*(s)}{s}$.) Since $\sum_i a_i^2 \le \frac{1}{2}$, choosing $s = t/(2\ell^*) = 2Lt$ gives

$$\Pr(\langle A, M\rangle \ge t) \le \exp\left(-\frac{t^2 L}{2} - \tilde{c}t^2 \sum_i a_i^2 \min\{1, (ta_i - s_*)^2\}\right).$$

To simplify the last term, note that because s_* is fixed, $x^2 \min\{1, (tx - s_*)^2\} = \min\{x^2, x^2(tx - s_*)^2\} \ge \Omega(\min\{x^2, (x/t - s_*)^2\})$, because if $(tx - s_*)^2 \le 1$ then $x^2 = \Theta(1/t^2)$ and so $x^2(tx - s_*)^2 = \Theta((x - s_*/t)^2)$. □

Corollary 41. *Let A be the adjacency matrix of a $\mathcal{G}(m, n)$ random graph with $m = p\binom{n}{2}$ and $p \ge \frac{1}{2}$. For c_* defined as in (8) and any $n^{-c_*} \ll t^k \ll 1$, conditioned on $\tau_k(A) \le p^k - t^k$ the* *following holds with high probability:* $\tilde{A} := A - \mathbb{E}A$ *has a unique (up to sign) unit eigenvector* v *with eigenvalue* $\lambda_n(\tilde{A})$*, and it satisfies*

$$\sum_{i,j} \min\{v_i^2 v_j^2, (v_i v_j - s_*/(tn))^2\} = o(1).$$

Proof. Uniqueness of the eigenvector follows from Theorem 1, which implies that the eigenvalue $\lambda_n(\tilde{A}) = -tn(1 - o(1))$ has multiplicity one. Also, Theorem 1 implies that

$$\Pr(\tau_k(A) \leq p^k - t^k) = \exp\left(-\frac{t^2 n^2 L}{2}(1 + o(1))\right),$$

and in order to show the claim it suffices to show that the probability of having a unit eigenvector v with eigenvalue $-tn(1 - o(1))$ and

$$\sum_{i,j} \min\{v_i^2 v_j^2, (v_i v_j - s_*/(tn))^2\} \geq \epsilon > 0 \tag{121}$$

is $o(\Pr(\tau_k(A) \leq p^k - t^k))$. For $\epsilon > 0$, let V_ϵ be the set of unit vectors v satisfying (121). First, note that for any fixed $v \in V_\epsilon$, Proposition 40 implies that

$$\Pr(\langle \tilde{A}, vv^T \rangle \leq -tn(1 - \delta)) \leq \exp\left(-\frac{t^2 n^2 L}{2}(1 - O(\delta) + \Omega(\epsilon))\right).$$

(Proposition 40 was written for matrices with i.i.d. entries, but we can apply it to $\tilde{A}$ by the standard trick of writing $\tilde{A}$ as a matrix with i.i.d. entries, conditioned on having a certain number of positive entries. The probability of the event we're conditioning on is $\Omega(1/n)$, and that extra factor of n can be absorbed in the $\exp(-\Omega(\epsilon)t^2 n)$ term.)

Now let $\mathcal{M}_{1,\delta} = \{vv^T : v \in V_\epsilon\}$, and by Lemma 15 there is a δ-net $\mathcal{N}$ for $\mathcal{M}_{1,\epsilon}$ of size at most $(C/\delta)^{Cn}$ (because we can start with an $(\delta/2)$-net of $\mathcal{M}_1$ and then project each element of that net onto $\mathcal{M}_{1,\epsilon}$, which gives an δ-net of $\mathcal{M}_{1,\epsilon}$). By Lemma 14 and a union bound, for any fixed $\delta > 0$ we have

$$\Pr\left(\inf_{M \in \mathcal{M}_{1,\epsilon}} \langle \tilde{A}, M \rangle \leq -tn(1 - \delta)(1 - 2\delta)\right) \leq \exp\left(-\frac{t^2 n^2 L}{2}(1 - O(\delta) + \Omega(\epsilon))\right),$$

because the $(C/\delta)^{Cn}$ term coming from the union bound can be absorbed in the $\exp(O(t^2 n^2 \delta))$ term. If δ is sufficiently small compared to ϵ, the probability bound above is asymptotically smaller than $\Pr(\tau_k(A) \leq p^k - t^k)$. Therefore, for every $\epsilon > 0$, conditioned on $\tau_k(A) \leq p^k - t^k$, with high probability the eigenvector of $\tilde{A}$ with eigenvalue $\lambda_n(\tilde{A})$ does not belong to V_ϵ. $\square$

We interpret Corollary 41 as saying that for the most-negative eigenvector of $\tilde{A}$, most of the ℓ^2 "mass" of $v_i v_j$ is concentrated near $s_* t^{-1} n^{-1}$. Our next task is to show that (after possibly changing the sign of v) v_i essentially takes two values: 0 and $s_*^{1/2} t^{-1/2} n^{-1/2}$. For notational convenience, we will adopt a different normalization: consider a vector w with the property that

$$\sum_{i,j} \min\{w_i^2 w_j^2, (w_i w_j - 1)^2\} \leq \epsilon |w|^4, \tag{122}$$

and we will show that (after possibly changing the sign of w) w is close to taking values zero and 1. To break the sign symmetry, we will assume without loss of generality that

$$\sum_{w_i<0} w_i^2 < \frac{1}{2}|w|^2. \tag{123}$$

Lemma 42. *If* (122) *and* (123) *hold then*

$$\sum_{w_i<0} w_i^2 \le 2\epsilon|w|^2, \tag{124}$$

$$\sum_{w_i\ge 1}(w_i-1)^2 \le \sqrt{\epsilon}|w|^2, \tag{125}$$

$$\sum_{\frac{1}{2}\le w_i\le 1}(w_i-1)^2 \le 3\sqrt{\epsilon}|w|^2, \text{ and} \tag{126}$$

$$\sum_{0\le w_i\le \frac{1}{2}} w_i^2 \le \sqrt{\epsilon}|w|^2. \tag{127}$$

In particular, if we define $\tilde{w}$ *by setting* $\tilde{w}_i \in \{0,1\}$*, whichever is closer to* w_i*, then* $|w-\tilde{w}|^2 \le 6\sqrt{\epsilon}|w|^2$.

Proof. If $w_i<0$ and $w_j>0$ then $w_iw_j<0$ and so $w_i^2w_j^2 \le (w_iw_j-1)^2$. Hence,

$$\sum_{w_i<0}\sum_{w_j>0} w_i^2w_j^2 \le \sum_{i,j}\min\{w_i^2w_j^2,(w_iw_j-1)^2\} \le \epsilon|w|^4.$$

If we define γ by $\sum_{w_i<0} w_i^2 = \gamma|w|^2$ then $\sum_{w_j<0} w_j^2 = (1-\gamma)|w|^2$ and so the equation above implies that $\gamma(1-\gamma)\le\epsilon$. By (123), $1-\gamma\ge\frac{1}{2}$ and so $\gamma\le 2\epsilon$. This proves (124).

To prove (125), define γ by $\sum_{w_i\ge1}(w_i-1)^2 = \gamma|w|^2$. Now, $w_i,w_j\ge 1$ implies that $(w_i-1)^2(w_j-1)^2 \le (w_iw_j-1)^2 \le w_i^2w_j^2$ It follows from (122) that

$$\gamma^2|w|^4 = \sum_{w_i,w_j\ge1}(w_i-1)^2(w_j-1)^2 = \sum_{w_i,w_j\ge1}(w_iw_j-1)^2 \le \sum_{i,j}\min\{w_i^2w_j^2,(w_iw_j-1)^2\} \le \epsilon|w|^4,$$

and it follows that $\gamma\le\sqrt{\epsilon}$.

To prove (126), define γ by $\sum_{\frac{1}{2}\le w_i\le1}(w_i-1)^2 = \gamma|w|^2$. If $\frac{1}{2}\le w_i,w_j\le 1$ then $w_i+w_j\ge 2w_iw_j$ and so $1-w_iw_j \ge (1-w_i)(1-w_j)$. Therefore,

$$\gamma^2|w|^4 = \sum_{\frac{1}{2}\le w_i,w_j\le 1}(w_i-1)^2(w_j-1)^2 \le \sum_{\frac{1}{2}\le w_i,w_j\le 1}(w_iw_j-1)^2.$$

Now, if $w_iw_j\ge\frac{1}{4}$ then $(w_iw_j-1)^2 \le 9\min\{w_i^2w_j^2,(w_iw_j-1)^2\}$, and so (122) implies that

$$\gamma^2|w|^4 \le 9\epsilon|w|^4.$$

To prove (127), define γ by $\sum_{0\le w_i\le\frac{1}{2}} w_i^2 = \gamma|w|^2$. If $0\le w_i,w_j\le\frac{1}{2}$ then $w_i^2w_j^2 = \min\{w_i^2w_j^2,(w_iw_j-1)^2\}$, and so

$$\gamma^2|w|^4 = \sum_{w_i,w_j\ge1} w_i^2w_j^2 \le \sum_{i,j}\min\{w_i^2w_j^2,(w_iw_j-1)^2\} \le \epsilon|w|^4,$$

and it follows that $\gamma \le \sqrt{\epsilon}$. ∎

We finally come to the proof of Theorem 3: let $\tilde{A}$ be the centered adjacency matrix of a $\mathcal{G}(n,m)$ random graph, and let v be a unit eigenvector with minimal eigenvalue; recall from Corollary 23 that with high probability v is unique, and that $w := n^{1/2}t^{1/2}s_*^{-1/2}v$ satisfies the condition (122) for some $\epsilon = o(1)$; for the rest of the proof, we will be working on this event. Without loss of generality (changing the sign if necessary) w also satisfies (123), and so Lemma 42 implies that there is a vector $\tilde{v}$ with $\tilde{v}_i \in \{0, n^{-1/2}t^{-1/2}s_*^{1/2}\}$ for all i, and $|\tilde{v} - v| = o(1)$. By Theorem 1 and Corollary 38, it follows that on this event,

$$\tilde{A} = -tn\tilde{v}\tilde{v}^T + R, \text{ where } \|R\|_{\mathrm{op}} = o(tn). \tag{128}$$

Let $U = \{i : \tilde{v}_i \neq 0\}$. Since $|\tilde{v}_i|^2 = 1 + o(1)$ and $v_i^2 \in \{0, n^{-1}t^{-1}s_*\}$, we must have $|U| = (1+o(1))nts_*^{-1}$. Now let V_1 and V_2 be any sets of vertices. Let 1_U denote the vector having $(1_U)_i = 1$ for $i \in U$, and $(1_U)_i = 0$ otherwise; and similarly for 1_{V_1} and 1_{V_2}. Then $\frac{1}{4}\langle A, (1_{V_1} + 1_{V_2})^{\otimes 2} - (1_{V_1} - 1_{V_2})^{\otimes 2}\rangle$ counts the number of edges between V_1 and V_2. Recalling that $A = \tilde{A} + p\mathbf{1} - pI$, on the event that (128) holds, the number of edges between V_1 and V_2 is

$$p|V_1||V_2| - (1+o(1))\frac{1}{4}(tn\langle \tilde{v}, 1_{V_1} + 1_{V_2}\rangle^2 - \langle \tilde{v}, 1_{V_1} - 1_{V_2}\rangle^2) + o(tn|1_{V_1} + 1_{V_2}|^2). \tag{129}$$

Recall that $\tilde{v} = n^{-1/2}t^{-1/2}s_*^{1/2}1_U$. Therefore, if $V_1, V_2 \subset U$ then $\langle \tilde{v}, 1_{V_1}\rangle = n^{-1/2}t^{-1/2}s_*^{1/2}|V_1|$ and similarly for V_2. Hence, the number of edges between V_1 and V_2 is

$$p|V_1||V_2| - (1+o(1))\frac{s_*}{4}((|V_1| + |V_2|)^2 - (|V_1| - |V_2|)^2) + o(|V_1 \cup V_2|^2) = (1-p)|V_1||V_2| + o(|U|^2).$$

When $V_1 \subset U^c$, we have $\langle \tilde{v}, 1_{V_1}\rangle = 0$ and so (129) implies that there are $p|V_1||V_2| + o(tn(|V_1| + |V_2|))$ edges between V_1 and V_2. This completes the proof of Theorem 3 in the case that either $V_1, V_2 \subset U$ or $V_1 \subset U^c$. To obtain the general case, we simply split V_i into $V_i \cap U$ and $V_i \cap U^c$.

References

[1] C. Goldschmidt, S. Griffiths and A. 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] S. Janson, Poisson approximation for large deviations, Random Structures & Algorithms **1** (1990) 221–229.

[3] S. Boucheron, G. Lugosi, P. Massart etal, Concentration inequalities using the entropy method, Ann. Prob. **31** (2003) 1583–1614.

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

[9] N. Cook and A. Dembo, Large deviations of subgraph counts for sparse Erdős-Rényi graphs, Adv. Math, **373** (2020) 107289.

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

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

[12] Y. Zhao, On the lower tail variational problem for random graphs, Combin. Probab. Comput. **26** (2017) 301–320.
[13] S. Bhattacharya and A. Dembo, Upper Tail For Homomorphism Counts in Constrained Sparse Random Graphs, arXiv:1909.03045
[14] B. Gunby, Upper Tails of Subgraph Counts in Sparse Regular Graphs, arXiv:2010.00658.
[15] E. Lubetzky and Y. Zhao, On replica symmetry of large deviations in random graphs, Random Structures Algorithms **47** (2015) 109–146.
[16] S. Janson and L. Warnke, The lower tail: Poisson approximation revisited, Random Structures and Algorithms **48** (2016) 219–246.
[17] S. Chatterjee, An introduction to large deviations for random graphs, Bull. Amer. Math. Soc. **53** 617–642.
[18] Svante Janson, *Orthogonal decompositions and functional limit theorems for random graph statistics*, American Mathematical Soc. 1994.
[19] M. Harel, F. Mousset and W. Samotij, Upper tails via high moments and entropic stability, arXiv:1904.08212.
[20] J. Neeman, C. Radin and L. Sadun, Phase transitions in finite random networks, J. Stat. Phys. **181** (2020) 30–328.
[21] C. Radin, K. Ren and L. Sadun, The asymptotics of large constrained graphs, J. Phys. A: Math. Theor. **47** (2014) 175001.
[22] R. Kenyon, C. Radin, K. Ren and L. Sadun, Multipodal structure and phase transitions in large constrained graphs, J. Stat. Phys. **168** (2017) 233–258.
[23] R. Kenyon, C. Radin, K. Ren and L. Sadun, Bipodal structure in oversaturated random graphs, Int. Math. Res. Notices, **2018** (2016) 1009–1044.
[24] C. Radin, K. Ren and L. Sadun, A symmetry breaking transition in the edge/triangle network model, Ann. Inst. H. Poincaré D **5** (2018) 251–286.
[25] R. Kenyon, C. Radin, K. Ren and L. Sadun, The phases of large networks with edge and triangle constraints, J. Phys. A: Math. Theor. **50** (2017) 435001.
[26] H. Koch, Vertex order in some large constrained random graphs, SIAM J. Math. Anal. **48** (2016) 2588–2601.
[27] S. Chatterjee and S.R.S. Varadhan, The large deviation principle for the Erdős-Rényi random graph, Eur. J. Combin. **32** (2011) 1000–1017.
[28] A. Dembo and E. Lubetzky, A large deviation principle for the Erdős-Rényi uniform graph, Electron. Commun. Probab. **23** (2018) 1–13.
[29] A. Guionnet and J. Husson, Large deviations for the largest eigenvalue of Rademacher matrices, Ann. Prob. **48** (2000) 1436–1465.
[30] F. Augeri, Large deviations principle for the largest eigenvalue of Wigner matrices without Gaussian tail, Electron. J. Probab. **21** (2016) 49 pp.
[31] B. Bhattacharya and S. Ganguly, Upper Tails for Edge Eigenvalues of Random Graphs, SIAM Journal on Discrete Mathematics **34** (2020) 1069–1083.
[32] F. Augeri, A. Guionnet and J. Husson, Large deviations for the largest eigenvalue of sub-Gaussian matrices, arXiv:1911.10591.
[33] A. Guionnet and O. Zeitouni, Concentration of the Spectral Measure for Large Matrices, Electron. Commun. Probab. **5** (2020) 119–136.
[34] T. Tao, *Topics in random matrix theory*, Americal Mathematical Soc. 2012.
[35] E. J. Candes and Y. Plan, *Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements*, IEEE Transactions on Information Theory **57.4** (2011) 2342–2359.
[36] R. D. Grigorieff, *A note on von Neumann's trace inequality*, Mathematische Nachrichten **151.1** (1991) 327–328.
[37] O. Kallenberg, *Foundations of Modern Probability, Second Edition*, Springer 2001.
[38] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, Springer 2010.

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