Probability Theory and Related Fields (2022) 184:613–679
https://doi.org/10.1007/s00440-022-01164-7

Large deviations for the largest eigenvalue of Gaussian networks with constant average degree

Shirshendu Ganguly[1] · **Kyeongsik Nam**[2]

Received: 6 May 2021 / Revised: 6 April 2022 / Accepted: 13 August 2022 /
Published online: 31 October 2022

Abstract
Large deviation behavior of the largest eigenvalue λ_1 of Wigner matrices including those arising from an Erdős-Rényi random graph $\mathcal{G}_{n,p}$ with i.i.d. random conductances on the edges has been the topic of considerable interest. However, despite several recent advances, not much is known when the underlying graph is sparse i.e., $p \to 0$, except the recent works (Bhattacharya et al., Ann Probab 49(4):1847–1885, 2021and Bhattacharya and Ganguly, SIAM J Discret Math, 2020) which consider the simpler case of the graph without additional edge weights. Under sufficiently general conditions on the conductance distribution, one expects the 'dense' behavior as long as the average degree np is at least logarithmic in n. In this article we focus on the case of constant average degree i.e., $p = \frac{d}{n}$ for some fixed $d > 0$ with standard Gaussian weights. Results in Bandeira and Van Handel (Ann Probab 44(4):2479–2506, 2016) about general non-homogeneous Gaussian matrices imply that in this regime λ_1 scales like $\sqrt{\log n}$. We prove the following results towards a precise understanding of the large deviation behavior in this setting.

1. (Upper tail probabilities and structure theorem): For $\delta > 0$, we pin down the exact exponent $\psi(\delta)$ such that

$$\mathbb{P}(\lambda_1 \geq \sqrt{2(1+\delta)\log n}) = n^{-\psi(\delta)+o(1)}.$$

   Further, we show that conditioned on the upper tail event, with high probability, a *unique maximal clique* emerges with a very precise δ dependent size (takes either *one* or *two* possible values) and the Gaussian weights are uniformly high

✉ Shirshendu Ganguly
sganguly@berkeley.edu

Kyeongsik Nam
ksnam@kaist.ac.kr

[1] Department of Statistics, Evans Hall, University of California, Berkeley, CA 94720, USA

[2] Department of Mathematical Sciences, KAIST, Daejeon, South Korea

in absolute value on the edges in the clique. Finally, we also prove an optimal localization result for the leading eigenvector, showing that it allocates most of its mass on the aforementioned clique which is spread uniformly across its vertices.

2. (Lower tail probabilities): The exact stretched exponential behavior

$$\mathbb{P}(\lambda_1 \le \sqrt{2(1-\delta)\log n}) = \exp\left(-n^{\ell(\delta)+o(1)}\right)$$

is also established.

As an immediate corollary, one obtains that λ_1 is typically $(1+o(1))\sqrt{2\log n}$, a result which surprisingly appears to be new. A key ingredient in our proofs is an extremal spectral theory for weighted graphs obtained by an ℓ_1-reduction of the standard ℓ_2-variational formulation of the largest eigenvalue via the classical Motzkin-Straus theorem [37], which could be of independent interest.

Mathematics Subject Classification 60F10 · 05C80 · 60B20 · 15A18

Contents

1 Introduction

Spectral statistics arising from random matrices and their asymptotic properties have been the subject of major investigations for several years. Fundamental observables of interest include the empirical spectral measure as well as edge/extreme eigenvalues. The study of such quantities began in the classical setting of the Gaussian unitary and orthogonal ensembles (GUE and GOE) where the entries are complex or real i.i.d. Gaussians up to symmetry constraints. These exactly solvable examples admit complicated but explicit joint densities for the eigenvalues which can be analyzed, albeit involving a lot of work, to pin down the precise behavior of several observables of interest.

The central phenomenon driving this article is the atypical behavior of the largest eigenvalue of a random matrix. This falls within the framework of large deviations which has attracted immense interest over the past two decades.

Perhaps not surprisingly, this was first investigated in the above mentioned exactly solvable cases [2, 3]. Subsequently, Bordenave and Caputo [17] considered empirical

distributions in Wigner matrices with entries with heavier tails where the large deviation is dictated by a relatively small number of large entries. This phenomenon was shown for the largest eigenvalue as well in [4].

Another set of random matrix models arise from random graphs, particularly the Erdős-Rényi graph $\mathcal{G}_{n,p}$ on n vertices with edge probability $p \in (0, 1)$. The literature on the study of such graphs is massive with a significant fraction devoted to the study of spectral properties. A long series of works established universality results for the bulk and edge of the spectrum in random graphs of average degree at least logarithmic in the graph size drawing similarities to the Gaussian counterparts (cf. [27, 28] and the references therein). For sparser graphs, however, including the case of constant average degree which is the focus of this article, progress has been relatively limited. Nonetheless, some notable accomplishments include the results in [1, 11, 12, 32] about the edge of the spectrum, as well as the results of [19] and [18], which studied continuity properties of the limiting spectral measure and a large deviation theory of the related local limits, respectively.

While large deviations theory for linear functions of independent random variables is by now classical (see [25]), recently a powerful theory of *non-linear* large deviations has been put forth, developed over several articles (some of which are reviewed below), which treats non-linear functions such as the spectral norm of a random matrix with i.i.d. entries.

Among the recent explosion of results around this, a series of works investigated spectral large deviations for $\mathcal{G}_{n,p}$, beginning with Chatterjee and Varadhan [23], where the authors proved a large deviation principle for the entire spectrum of $\mathcal{G}_{n,p}$ at scale np, building on their seminal work [22], in the case where p is fixed and does not depend on n (dense case). However, the sparse case where $p = p(n) \to 0$ was left completely open until a major breakthrough was made by Chatterjee and Dembo [21]. This led to considerable progress in developing the theory of large deviations for various functionals of interest for sparse random graphs [5, 7, 10, 26, 42].

Closest, in spirit, to the results of this paper are two recent works that we describe next. Via a refined understanding of cycle counts in $\mathcal{G}_{n,p}$ which was obtained in [5, 9, 15, 22, 24, 31, 35, 36], one can deduce large deviation properties for eigenvalues using the trace method and this was carried out in [14]. However such arguments only extended to p going to zero at a rate slower than $1/\sqrt{n}$, since cycle statistics fail to encode information about the spectral norm for sparser graphs. Such sparser graphs were treated more recently in [13], where the first named author along with Bhaswar Bhattacharya and Sohom Bhattacharya analyzed the large deviations behavior for the spectral edge for sparse $\mathcal{G}_{n,p}$ in the *entire* "localized regime" when

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

where the extreme eigenvalues are governed by high degree vertices. This notably includes the well studied example of constant average degree.

At the law of large numbers level, as established in [27, 28, 32], $\lambda_1 = (1 + o(1)) \max(d_1, np)$ where d_1 denotes the maximum degree of the random graph. Con-

sequently λ_1 exhibits a transition at $np = \sqrt{\frac{\log n}{\log\log n}}$, where the largest eigenvalue begins to be governed by the largest degree. A similar phenomenon reflecting this transition for large deviations was established across the papers [13, 14]. In the case of Gaussian ensembles, although a precise result does not appear in the literature to the best of the authors' knowledge, it is expected that the dense behavior extends to the case of the average degree being logarithmic in n (an analogous result for Wigner matrices with bounded entries, which is more comparable to the setting of random graphs, was established in [40]). Beyond this, as the graph becomes sparser, a different behavior is expected to emerge.

This motivates the present work where we obtain a very precise understanding, of the case of constant average degree, i.e., $p = \frac{d}{n}$, arguably the most interesting sparse case because of its connections to various models of statistical mechanics.

Recently, [6, 30] have explored universality of large deviations behavior for the largest eigenvalue for a Wigner matrix with i.i.d sub-Gaussian entries relying on considering appropriate tilts of the original measures and analyzing the associated spherical integrals. Interesting behavior is shown to emerge in [6] when the sub-Gaussian tails are not sharp. Perhaps the most interesting examples in this class are sparsified Gaussian matrices whose entries are obtained by multiplying a Gaussian variable with an independent Bernoulli random variable with mean p. However, the methods has been shown to work only in the 'dense' case of constant p where the typical behavior is still the same as when $p = 1$, leaving the sparse regime $p \ll 1$ completely open, calling for new methods to treat sparser graphs.

Also relevant to this paper is a different line of research, which, motivated by viewing a random matrix as a random linear operator, considers 'non-homogeneous' matrices. The most well studied example is a Gaussian matrix where the variance varies from entry to entry. In this general setting, even the leading order behavior for the spectral norm is far from obvious and requires a much more refined understanding beyond the concentration of measure bounds obtained as a consequence of the non-commutative Khintchine inequality. A beautiful conjecture posed by Latala [33] related to an earlier result of Seginer [39] states that the expected spectral norm for such non-homogeneous Gaussian matrices, is up to constants the expectation of the maximum ℓ_2 norm of a row and a column, and after a series of impressive accomplishments [8, 41], the conjecture was finally settled in the beautiful work [34].

Note that sparse Wigner matrices, quenching on the sparsity, falls in the above framework where the variance of each entry is 0 or 1. It is worth mentioning that while the dependence on n in the leading order behavior is pinned down in the above mentioned works, the techniques are not sharp enough to unearth finer properties such as the exact constant multiplicative pre-factor.

We now move on to the statements of the main theorems after setting up some basic notations.

1.1 Setup and main results

We will denote by $\mathscr{G}_n$ the set of all simple, undirected networks on n vertices labelled $[n] := \{1, 2, \ldots, n\}$ i.e., simple graphs with a conductance value on each edge. For

$G \in \mathcal{G}_n$, denote by $A(G) = (a_{ij})_{1\le i,j\le n}$, the adjacency matrix of G, that is a_{ij} is the conductance associated to the edge (i, j) if the latter is an edge in G, and 0 otherwise. Thus graphs are trivially encoded as networks where the entries of A are 0 or 1. For $F \in \mathcal{G}_n$, since $A(F)$ is a self-adjoint matrix, denote by $\lambda_1(F) \ge \lambda_2(F) \ge \cdots \ge \lambda_n(F)$ its eigenvalues in non-increasing order, and let $\|F\|_{\text{op}} := \|A(F)\|_{\text{op}} = \max\{|\lambda_1(F)|, |\lambda_n(F)|\}$ be the operator norm of A. Throughout most of the paper we will be concerned with $\lambda_1(F)$ and for notational brevity we will often drop the subscript to denote the same.

In this paper we are interested in the sparse Erdős-Rényi random graph $\mathcal{G}_{n,p}$, where $p = \frac{d}{n}$ for some $d > 0$ which does not depend on n. We will denote by X the random adjacency matrix associated to it. Thus for all $1 \le i < j \le n$, $X_{i,j}$ is an independent Bernoulli random variable with mean p, and $X_{ii} = 0$ for all i. Let Y be a symmetric matrix given by $Y_{ij} \sim N(0, 1)$ for $i \le j$. The matrix of interest for us is $Z = X \odot Y$, i.e., $Z_{ij} = X_{ij}Y_{ij}$. Note that since $X_{ii} = 0$ for all i, in this setup, the diagonal entries of Y do not play any role.

Let $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n$ be eigenvalues of the matrix Z. As a consequence of the already referred to work on the behavior of the spectral norm of general inhomogeneous Gaussian matrices [8], it follows that

$$\mathbb{E}(\lambda_1) \approx \sqrt{\log n}. \tag{2}$$

One also obtains concentration around $\mathbb{E}(\lambda_1)$ using standard Gaussian techniques, see e.g. [8, Corollary 3.9]. However so far, the methods have not been able to obtain a sharper understanding including the precise constant in front of $\sqrt{\log n}$ which we deduce as a simple corollary of our main theorems. We now move on to the exact statements of the results in this paper.

Theorem 1.1 (Upper tail probabilities). *For $\delta > 0$, define a function $\phi_\delta : \mathbb{N}_{\ge 2} \to \mathbb{R}$*[1] *by*

$$\phi_\delta(k) := \frac{k(k-3)}{2} + \frac{1+\delta}{2}\frac{k}{k-1} \tag{3}$$

and $\psi(\delta) := \min_{k\in\mathbb{N}_{\ge 2}} \phi_\delta(k)$. Then,

$$\lim_{n\to\infty} -\frac{1}{\log n} \log \mathbb{P}(\lambda_1 \ge \sqrt{2(1+\delta)\log n}) = \psi(\delta). \tag{4}$$

Remark 1.2 (Infinite phase transition in upper tail). The rate function given by (4) is a continuous piecewise linear function with infinitely many pieces which we now describe in detail. Since we will only be concerned about the arg min restricted to integers larger than 1, we consider momentarily $\phi_\delta(x) = \frac{x(x-3)}{2} + \frac{1+\delta}{2}\frac{x}{x-1}$ as a

[1] $\mathbb{N}$ will be used to denote the set of natural numbers, and $\mathbb{N}_{\ge k}$ to denote all the natural numbers bigger equal to k.

function of real numbers greater than one and notice that,

$$\phi_\delta'(x) = x - \frac{3}{2} - \frac{1+\delta}{2}\frac{1}{(x-1)^2}. \tag{5}$$

Thus $\phi_\delta(x)$ is a strictly convex function. Let $\mathcal{M}(\delta) = \{\arg\min_{k\geq 2} \phi_\delta(k)\}$ be the set of minimizers of $\phi_\delta(\cdot)$. By the strict convexity of $\phi_\delta(\cdot)$, $\mathcal{M}(\delta)$ is at most of size 2 containing either a single element or two consecutive integers. Precisely, denoting by $x(\delta) > 1$, the unique solution to $\phi_\delta'(x) = 0$, any element in $\mathcal{M}(\delta)$ is either $\lfloor x(\delta)\rfloor$ or $\lceil x(\delta)\rceil$. Now the values of δ for which $\mathcal{M}(\delta)$ is of size two forms a discrete set. That is, there exists $0 = \delta_1 < \delta_2 < \delta_3 < \cdots$ such that the following holds: for any positive integer $k \geq 2$, (δ_{k-1}, δ_k) is the collection of δ such that $\mathcal{M}(\delta) = \{k\}$ and δ_k is the unique δ such that $\mathcal{M}(\delta) = \{k, k+1\}$. To see this, since $\delta \mapsto x(\delta)$ is strictly increasing, it suffices to verify that the situation $\delta_1 < \delta_2$, $\phi_{\delta_1}(k+1) \leq \phi_{\delta_1}(k)$ and $\phi_{\delta_2}(k) \leq \phi_{\delta_2}(k+1)$ never occurs. Observe that the contrary implies

$$\phi_{\delta_1}(k+1) \leq \phi_{\delta_1}(k) \leq \phi_{\delta_2}(k) \leq \phi_{\delta_2}(k+1).$$

By (5), $\phi_{\delta_1}'(x) > \phi_{\delta_2}'(x)$, which contradicts the above.

Hence, for $\delta \in [\delta_{k-1}, \delta_k]$,

$$\psi(\delta) = \frac{1+\delta}{2}\frac{k}{k-1} + \frac{k(k-3)}{2},$$

which is a linear function in $\delta \in [\delta_{k-1}, \delta_k]$ for any fixed $k \geq 2$. This implies that $\psi(\delta)$ is a continuous piecewise linear function.

Also by a simple algebra, it follows from (5) that

$$\left(\frac{1+\delta}{2}\right)^{1/3} + 1 < x(\delta) < \left(\frac{1+\delta}{2}\right)^{1/3} + \frac{3}{2}.$$

Since

$$\phi_\delta\left(\left(\frac{1+\delta}{2}\right)^{1/3}\right) = \frac{1}{2}\delta + \frac{3}{2^{5/3}}\delta^{2/3} + O(\delta^{1/3}),$$

we obtain

$$\psi(\delta) = \frac{1}{2}\delta + \frac{3}{2^{5/3}}\delta^{2/3} + O(\delta^{1/3}) \quad \text{as} \quad \delta \to \infty, \tag{6}$$

where $O(\delta^{1/3})$ is a quantity bounded by $C\delta^{1/3}$ for some absolute constant $C > 0$. Plugging this into (4), one thus obtains the following asymptotic behavior of the upper

tail probabilities[2]

$$\mathbb{P}(\lambda_1 \geq \sqrt{2(1+\delta)\log n}) = n^{-(\frac{1}{2}\delta + \frac{3}{2^{5/3}}\delta^{2/3} + O(\delta^{1/3}))} \text{ for large } \delta > 0, \text{ and,} \quad (7)$$

$$\mathbb{P}(\lambda_1 \geq \sqrt{2(1+\delta)\log n}) = n^{-\delta + o(1)} \text{ for small } \delta > 0. \quad (8)$$

Remark 1.3 (Comparison with maximum of i.i.d. Gaussians). As the reader possibly already notices, for small δ, the behavior in (8) is the same as that for the maximum of n many standard Gaussian variables. The reason for this will be discussed in the idea of proofs section.

Having established the sharp order of the tail probabilities, we now investigate the structural behavior conditioned on the upper tail event $\mathcal{U}_\delta := \{\lambda_1 \geq \sqrt{2(1+\delta)\log n}\}$. It is worth emphasizing that though in large deviations one often has a guess for the dominant mechanism guiding rare events, a precise theorem verifying the same is seldom obtained, usually since the latter typically relies on refined probability estimates which are usually difficult to establish.

However, fortunately, this is a rare occasion where the arguments do permit us to prove three results which provide a rather complete and precise understanding of the underlying structural effect of large deviations.

The first result shows the existence of a clique of a very precise δ dependent size establishing a sharp concentration for the maximal clique size conditioned on $\mathcal{U}_\delta$. For any graph $G \in \mathcal{G}_n$, let k_G be the size of a maximal clique K_G in G. Recall the definition of $\mathcal{M}(\delta)$ and let $h(\delta)$ be the smallest element of $\mathcal{M}(\delta)$. By Remark 1.2,

$$\left| h(\delta) - \left(\frac{1+\delta}{2}\right)^{1/3} - 1 \right| \leq 2. \quad (9)$$

Theorem 1.4 (Structure theorem). *For any δ with $h(\delta) \geq 4$, i.e., $\delta > \delta_3$ (recall the definition of δ_k from Remark 1.2),*

$$\lim_{n\to\infty} \mathbb{P}\Big(k_X \in \mathcal{M}(\delta) \mid \mathcal{U}_\delta\Big) = 1. \quad (10)$$

Furthermore, with conditional probability tending to one, K_X is unique and any clique of size at least 4 is a subset of K_X.

Note that the above statement in particular implies that the largest clique outside K_X is a triangle whose occurrence has constant probability. Thus the above result proves a two point concentration for the maximal clique size and for values of δ such that $\mathcal{M}(\delta)$ only contains $h(\delta)$, it implies a one point concentration.

Remark 1.5 Although the statement only accounts for $\delta > \delta_3$, with a bit more work, albeit somewhat technical, one may also prove (10) for $\delta > \delta_2$ (owing to the technical

[2] Throughout the paper, $o(1)$ will be used to denote functions of n that tend to 0 as n tends to infinity. However we will also need to deal with quantities that go to zero as δ converges to infinity, which would be denoted by $o_\delta(1)$.

nature of the argument we refrain from stating and proving a formal statement). On the other hand, when $\delta \le \delta_3$ (i.e. $h(\delta) \le 3$), the uniqueness part of the maximal clique K_X in Theorem 1.4 becomes false since with only constant additional probability cost, $\mathcal{G}_{n,\frac{d}{n}}$ possesses $O(1)$ many additional triangles.

In the case $\delta \le \delta_2$ (i.e., the minimum element in $\mathcal{M}(\delta)$ is 2), it is not hard to see that even the statement in (10) is false. To see this, note that while the large deviations in this case is dictated by a large weight on a single edge, the graph typically possesses $O(1)$ many triangles which should continue to persist even in the large deviations regime.

Now, this would have been immediate if the largest eigenvalue and the underlying graph X were positively correlated. While that is not quite the case (since the edge weights can take negative values), one can indeed show a *weakly* positive correlation, i.e., conditioned on $\mathcal{U}_\delta$, X stochastically dominates $\mathcal{G}_{n,\frac{d}{2n}}$ (where the edge density is half of the original density). This can be obtained by observing that the conditional probability of an edge being present conditional on $\mathcal{U}_\delta$, the remainder of the graph and the associated edge weights, is at least $\frac{d}{2n}$. The desired claim now follows from the fact that the probability that $\mathcal{G}_{n,\frac{d}{2n}}$ contains a triangle is uniformly bounded from below by a positive number.

Our next result asserts that most of the contribution to the spectral norm comes from K_X, with the Gaussians along the edges of the latter being uniformly high in absolute value.

Theorem 1.6 (Uniformly high Gaussian values). *There exists $\zeta = \zeta(\kappa) > 0$ with $\lim_{\kappa\to 0}\zeta = 0$ such that the following holds. For $\kappa > 0$, for δ large enough, with probability (conditional on $\mathcal{U}_\delta$) going to 1, there exists $T \subset K_X$ such that $|T| \ge (1-\kappa)h(\delta)$ and*

$$\frac{1}{h(\delta)^2}\sum_{i\neq j, i,j\in T}\left||Z_{ij}| - \frac{1}{h(\delta)}\sqrt{2(1+\delta)\log n}\right| \le \frac{\zeta}{h(\delta)}\sqrt{2(1+\delta)\log n}. \tag{11}$$

Even though in the statement δ is chosen large enough as a function of κ, the proof will in fact give a quantitative, albeit technical, bound for all large δ and small κ which can then be simplified into the form of the statement of the theorem by choosing δ dependent on κ.

Since the maximal clique K_X has size $h(\delta)$ or $h(\delta)+1$ with probability going to 1 (conditional on $\mathcal{U}_\delta$), the above theorem shows that the Gaussian values Z_{ij} on K_X are uniformly high in absolute value and close to $\frac{1}{h(\delta)}\sqrt{2(1+\delta)\log n}$ in the ℓ_1 sense.

Our final structural result is an optimal localization statement about the leading eigenvector.

Theorem 1.7 (Optimal localization of eigenvector). *Let $\boldsymbol{v} = (v_1, \cdots, v_n)$ be the top eigenvector with $\|\boldsymbol{v}\|_2 = 1$ and consider the unique maximal clique K_X and its size k_X from Theorem 1.4. For $\kappa > 0$, define the events*

$$\mathcal{A}_1 := \left\{\sum_{i\in K_X} v_i^2 \ge 1-\kappa\right\}$$

and

$$\mathcal{A}_2 = \Big\{ \frac{1}{k_X} \sum_{i \in K_X} \Big(v_i^2 - \frac{1}{k_X} \Big)^2 \le \frac{40\kappa}{k_X^2} \Big\}.$$

Then, for sufficiently large $\delta > 0$,

$$\lim_{n \to \infty} \mathbb{P}(\mathcal{A}_1 \cap \mathcal{A}_2 \mid \mathcal{U}_\delta) = 1. \tag{12}$$

Thus the above theorem says, for any $\kappa > 0$, for all large enough n, conditioned on $\mathcal{U}_\delta$, the leading eigenvector distributes at least $1 - \kappa$ mass on K_X almost uniformly.

Note that the last two theorems do not claim anything about the sign of the entries of the eigenvector or the Gaussian values. This is since switching the signs of the entries of the largest eigenvector arbitrarily and accordingly changing the signs of the Gaussians yields the same quadratic form.

Having stated our results concerning upper tail deviations, the next result pins down the lower tail large deviation probability.

Theorem 1.8 (Lower tail probabilities). *For any* $0 < \delta < 1$,

$$\lim_{n \to \infty} \frac{1}{\log n} \Big(\log \log \frac{1}{\mathbb{P}(\lambda_1 \le \sqrt{2(1-\delta)\log n})} \Big) = \delta. \tag{13}$$

As an immediate corollary of Theorems 1.1 and 1.8, one obtains the following *'law of large numbers'* behavior which we were surprised to not be able to locate in the literature.

Corollary 1.9 *We have*

$$\lim_{n \to \infty} \frac{\lambda_1}{\sqrt{\log n}} = \sqrt{2}$$

in probability.

We conclude this discussion by remarking that although in principle our techniques may be used to analyze a wider subset of the parameter space, we have, for concreteness and aesthetic considerations, chosen to simply focus on the case of constant average degree.

1.2 Organization of the article

In Sect. 2 we provide a detailed account of the keys ideas driving the proofs. In Sect. 3, we state and prove the key Proposition 3.1 obtaining a bound on the spectral norm in terms of the Frobenius norm for weighted graphs. The rest of the paper focuses on the proofs of Theorem 1.1 (Sects. 4, 5), Theorem 1.4 in Sect. 6, Theorem 1.7 in Sect. 7, Theorem 1.6 in Sect. 8 and Theorem 1.8 in Sect. 9 respectively. Certain straightforward but technical estimates are proved in the "Appendix".

2 Key ideas of the proofs

In this section, we provide a sketch of the arguments in the proofs of our main results and end with a brief comparison of the approach in this paper to related existing work in the literature.

Upper tail lower bound: This is straightforward. The strategy is to plant a clique of an appropriate size ($\arg\max_k \phi_\delta(k)$) and have high valued Gaussians on all the clique edges, i.e., at least $\frac{\sqrt{2(1+\delta)\log(n)}}{k-1}$. The probability of a clique of size $k \geq 3$ appearing is up to constants $n^{k-\binom{k}{2}}$ (the proof follows by a second moment argument) while the probability of having high Gaussians is

$$\mathbb{P}\Big(Y_{ij} \geq \frac{\sqrt{2(1+\delta)\log(n)}}{k-1},\ \forall\ 1 \leq i < j \leq k\Big) \geq \Big(\frac{C}{\sqrt{\log n}} n^{-\frac{1+\delta}{(k-1)^2}}\Big)^{\binom{k}{2}},$$

where the right hand side follows from standard Gaussian tail bounds (see (24) later). Thus the total cost at the polynomial scale is $n^{k-\binom{k}{2}} n^{-\frac{1+\delta}{(k-1)^2}\binom{k}{2}}$. Observe that the exponent is precisely $-\phi_\delta(k)$. When $k = 2$, one should view it slightly differently however, since $k - \binom{k}{2} = 1 > 0$. Namely, there are order $n^{k-\binom{k}{2}} = n$ many edges and hence the probability that there exists a Gaussian of value at $\sqrt{2(1+\delta)\log n}$ is $nn^{-(1+\delta)} = n^{-\delta} = n^{-\phi_\delta(2)}$. Finally, optimizing over k yields the bound $n^{-\psi(\delta)}$.

It is worth noticing the contrasting behavior in the absence of the Gaussian variables, where in [13] it was shown that large deviations for the largest eigenvalue is guided by the large deviations for the maximum degree and not by appearance of a clique.

Upper tail upper bound: This is the most difficult among the four bounds and a significant part of the work goes into proving this. The first step is to make the underlying graph sparser by only focusing on the Gaussians with a large enough value. This is a trick that has appeared in some form in previous works (see e.g. [13, 17]) and the most delicate part of our arguments goes into analyzing the sparser graph obtained by this method.

Broadly speaking, the reason for the sparsification is two-fold. a) It is much harder for the graph restricted to small Gaussian values to have a high spectral norm, and so for our purposes we will treat that component as *spectrally negligible*. This relies on recent results from [13]. b) The graph restricted to high Gaussian values is much sparser and hence admits greater shattering into smaller components whose sizes we can control; since eigenvalues of different components do not interact with each other, this will be particularly convenient, albeit modulo a refined analysis of the connectivity structure of the latter.

Proceeding to implement this strategy, decompose the Gaussian random variables Y_{ij} as

$$Y_{ij} = Y_{ij}^{(1)} + Y_{ij}^{(2)},$$

where $Y_{ij}^{(1)} = Y_{ij}\mathbb{1}_{|Y_{ij}|>\sqrt{\varepsilon \log\log n}}$ and similarly $Y_{ij}^{(2)} = Y_{ij}\mathbb{1}_{|Y_{ij}|\le\sqrt{\varepsilon \log\log n}}$. Thus, we can write the matrix Z as $Z^{(1)} + Z^{(2)}$ with

$$Z_{ij}^{(1)} = X_{ij}Y_{ij}^{(1)}, \quad Z_{ij}^{(2)} = X_{ij}Y_{ij}^{(2)}, \tag{14}$$

and similarly $X = X^{(1)} + X^{(2)}$ i.e., $X_{ij}^{(1)} = X_{ij}\mathbb{1}_{|Y_{ij}|>\sqrt{\varepsilon \log\log n}}$. We next prove an upper bound on the probability that $Z^{(2)}$ has high spectral norm which is much smaller than that for Z which implies that the spectral behavior of Z even under large deviations is dictated by that of $Z^{(1)}$. The choice of the truncation threshold is governed by the fact that the typical spectral norm of $\mathcal{G}_{n,\frac{d}{n}}$ is of order $\sqrt{\frac{\log n}{\log\log n}}$ which in itself is a consequence of the fact that the maximum degree is of order $\frac{\log n}{\log\log n}$. Sharp large deviations behavior for eigenvalues of sparse random graphs was recently established in the already mentioned work [13] which we use to make this step precise.

This allows one to focus simply on $Z^{(1)}$ or the underlying graph $X^{(1)}$, conditioning on which makes the spectral behavior of the individual connected components independent. Further note that the edge weights are independent Gaussian variables each of which is conditioned to be at least $\sqrt{\varepsilon}\sqrt{\log\log n}$.

Let $C_1, \cdots, C_k$ be the connected components of $X^{(1)}$. At this point, denoting the network Z restricted to C_ℓ by Z_ℓ, we relate $\|Z_\ell\|_{\text{op}}$ to its Frobenius norm $\|Z_\ell\|_F$. The trivial bound $\|Z_\ell\|_{\text{op}} \le \|Z_\ell\|_F$ is easy to see. The next idea which is the key one in this paper relies on the following sharp improvement over the above. Namely, we show that if k_ℓ is the size of the maximal clique in Z_ℓ, then

$$\|Z_\ell\|_{\text{op}}^2 \le \frac{k_\ell - 1}{k_\ell}\|Z_\ell\|_F^2. \tag{15}$$

The proof of the above relies on reducing the standard ℓ_2 variational problem for the spectral norm to an ℓ_1 version, which allows us to use the classical work of Motzkin and Straus [37]. This leads to a bound of the form

$$\mathbb{P}(\|Z_\ell\|_{\text{op}}^2 \ge 2(1+\delta)\log n) \le \mathbb{P}\Big(\|Z_\ell\|_F^2 \ge \frac{k_\ell}{k_\ell - 1}2(1+\delta)\log n\Big). \tag{16}$$

Now quenching the graph X, the random variable $\|Z_\ell\|_F^2$ can be viewed at first glance as a chi-squared random variable with degrees of freedom given by the component size $|E(C_\ell)|$. Now as long as $|E(C_\ell)|$ is $o(\log n)$, the degree of freedom does not affect the latter probability in its leading order behavior and it behaves as the square of a single Gaussian. This is what justifies the sparsification step mentioned at the outset which ensures that $|C_\ell| = O_\varepsilon(\frac{\log n}{\log\log n})$ which along with the tree like behavior of C_ℓ implies $|E(C_\ell)| = O_\varepsilon(\frac{\log n}{\log\log n})$ as well (Here $O_\varepsilon(\cdot)$ is the standard notation denoting that the implicit constant is a function of ε.)

However there is one crucial subtlety that we have overlooked so far. Namely, $\|Z_\ell\|_F^2$ is not simply a chi-squared random variable but instead is a sum of squares of independent Gaussian variables each conditioned to have an absolute value at least

$\sqrt{\varepsilon \log\log n}$. This makes the tail heavier by the exact amount which on interacting with the ε dependence in the size of C_ℓ begins to affect the leading order probability. Thus, unfortunately, the above strategy ends up not quite working.

To address this we further rely on the fact that C_ℓ is almost tree-like and has a bounded number of 'tree-excess edges' with high probability and revise our strategy in the following way. Consider the eigenvector v corresponding to the largest eigenvalue $\lambda(\ell) := \lambda_1(C_\ell)$. Thus we know $v^\top Z_\ell v = \lambda(\ell)$.

The key idea now is to split the vertices of C_ℓ, according to high and low values of v. We first show that it is much more costly for the Frobenius norm to be high on the subgraph induced by the low values of v. This is where the tree like property is crucially used as well.

Thus we focus only on the $O(1)$ vertices supporting high v values and since the maximum degree is $O(\frac{\log n}{\log\log n})$ (*without* an ε dependence in the constant), the strategy originally outlined can be made to work for the subgraph induced by these vertices.

While the next three proofs are rather technically involved, here we simply review the high level strategies involved.

Emergence of a unique maximal clique: The above proofs imply that the graph $X^{(1)}$ under $\mathcal{U}_\delta$ contains a clique $K_{X^{(1)}}$ whose size is sharply concentrated on $\mathcal{M}(\delta)$ (where the latter appearing in the statement of Theorem 1.4 denotes the set of minimizers of $\phi_\delta(\cdot)$). It also follows that $K_{X^{(1)}}$ is unique. We then show that on account of sparsity, superimposing $X^{(2)}$ on $X^{(1)}$ does not alter this. Particularly convenient is the fact that conditional on $X^{(1)}$, the spectral behavior of $Z^{(1)}$ and the random graph $X^{(2)}$ are independent. However making this precise is delicate and is one of the most technical parts of the paper, relying on a rather refined understanding of the graph $X^{(1)}$ under the large deviation behavior of $\lambda(Z^{(1)})$ Such understanding also allows us to show that there does not exist any other clique in X of size at least 4 which is not contained in K_X.

Localization of the leading eigenvector: The proof of this is reliant on the fact that (15) is sharp only when the leading eigenvector is supported on the maximal clique K_X. We prove a quantitative version of this fact showing that significant mass away from the clique results in a deteriorated form of (15) which then makes $\mathcal{U}_\delta$ much more costly than the already proven lower bound for its probability. Further a similar approach is used to prove the desired flatness of the vector on K_X.

Flatness of the Gaussian values on the maximal clique. Using the previous structural result about the leading eigenvector $v = (v_1, v_2, \ldots, v_n)$, we consider the set $T \subset K_X$ such that $|v_i| \approx \frac{1}{k_X}$ for all $i \in T$ (we don't make the meaning of $\approx$ precise) The previous results guarantee that, conditional on $\mathcal{U}_\delta$, $|T| \geq (1-\kappa)k_X$ and $|k_X - h(\delta)| \leq 1$. Firstly showing that the spectral contribution from the edges incident on T^c is negligible, it follows that the quadratic form $v^\top Z v \approx v_T^\top Z_T v_T$ where v_T and Z_T are the restrictions to the subgraph induced on T. Now owing to the flatness of v on T (and this is why we work with T and not K_X), it follows that

$$v_T^\top Z_T v_T \leq 2\frac{(1+o_\delta(1))}{h(\delta)}\|Z_T\|_1,$$

where the ℓ_p norm Z_T is defined by

$$\|Z_T\|_p := \Big(\sum_{i<j, i,j\in T} |Z_{ij}|^p \Big)^{1/p}.$$

Using this and the fact that $|k_X - h(\delta)| \le 1$ with high probability, we obtain the bound

$$\|Z_T\|_1 \approx \frac{1}{2} h(\delta) \sqrt{2(1+\delta')\log n}.$$

In fact the above argument only implies a lower bound, while the upper bound follows from the following sharp bound on the ℓ_2 norm which is a consequence of previous arguments (e.g. (16)).

$$\|Z_T\|_2^2 \approx (1 + o_\delta(1))(1+\delta)\log n.$$

Using the above two bounds, one can conclude the statement of the theorem in a straightforward fashion.

Lower tail: The upper bound can be obtained simply by a comparison with the maximum of $O(n)$ many independent Gaussians.

For the lower bound, $Z^{(2)}$ can still be considered spectrally negligible, while for $Z^{(1)}$, conditioning on $X^{(1)}$ being 'nice', with none of the components being too large while also having at most bounded tree excess we use the results about the upper tail to upper bound the probability that for any connected component C_ℓ, $\lambda(\ell) \ge \sqrt{2(1-\delta)\log n}$ or in other words lower bound $\mathbb{P}(\lambda(\ell) \le \sqrt{2(1-\delta)\log n})$ where $\lambda(\ell) := \lambda_1(C_\ell)$. Since $\lambda_1(Z) = \max_\ell(\lambda(\ell))$ and, conditioning on the graph makes $\lambda(\ell)$ across different values of ℓ independent, the result follows in a straightforward fashion.

2.1 Comparison to past works

We end with a brief discussion on past results on large deviation of spectral statistics for various examples of random matrices to contrast them with the statements and the proof method in this paper.

As mentioned in the introduction, the earliest results for Gaussian ensembles [2, 3] relied on the exact form of the joint density that the eigenvalues admit. In [17] empirical distributions of Wigner matrices with entries with heavier tails than Gaussians were considered. The authors used a thresholding argument (which, recall, is also the first step in our approach) to decompose the matrix as a sum of a typical matrix and a sparse matrix of large entries, thereby expressing the limiting spectral measure as a free convolution of the semi-circle law (from the typical part) and the spectral measure of a Sofic measure on random networks (from the sparse part). A similar strategy was carried out later for the largest eigenvalue in [4]. Recently, [6, 30] have explored universality of large deviations behavior for the largest eigenvalue for a Wigner matrix with i.i.d sub-Gaussian entries relying on considering appropriate tilts of the original measures and analyzing the associated spherical integrals.

More closer in spirit to our present work is the work on the spectral norm of the Erdős-Rényi graph $\mathcal{G}_{n,p}$ when $p \to 0$. Using connection to cycle statistics, [14] proved large deviation behavior for the largest eigenvalue as long as p is going to zero at a rate slower than $1/\sqrt{n}$. For sparser graphs including the case of constant average degree, in absence of any edge weights, [13] showed that the large deviation behavior of the extreme eigenvalues are governed by high degree vertices by decomposing the graph into a low degree part and a disjoint union of high degree vertices.

As we see in the present work, addition of Gaussian weights has a drastic effect on the nature of large deviations, with cliques with high Gaussian values leading to large spectral norm in place of high degree vertices. Our analysis is also significantly different and complicated compared to [13] where in the latter, the decomposition simply involved picking out the high degree vertices while in the present work, one has to analyze, to a rather refined detail, the various connected components of the sparse graph obtained as a result of the thresholding step in (40).

Finally, it is worth reiterating that the structure theorems obtained in this paper stand sharply in contrast to past results of a similar flavor due to the unusual degree of their precision.

3 Spectral theory of weighted graphs

As outlined in Sect. 2, a key ingredient in our proofs is a new deterministic bound on the spectral norm in terms of the Frobenius norm by an $\ell_2 \to \ell_1$ reduction allowing us to use the classical Motzkin-Strauss theorem. Though this is independently interesting, the proofs are somewhat technical and the reader only interested in the large deviations aspect, at first read can simply treat this result as an input in the proof of Theorem 1.1.

3.1 Spectral norm and Frobenius norm

For a Hermitian matrix A of size $n \times n$, let $\lambda_1 \geq \cdots \geq \lambda_n$ be the eigenvalues in a non-increasing order. Then, we have

$$\mathrm{tr}(A^k) = \lambda_1^k + \cdots + \lambda_n^k,$$

which immediately implies that for any even positive integer k,

$$\lambda_1^k \leq \mathrm{tr}(A^k) \leq n\lambda_1^k. \tag{17}$$

We denote by $\|A\|_F$, the Frobenius norm of the matrix A:

$$\|A\|_F := (\mathrm{tr}(A^2))^{1/2} = \Big(\sum_{1 \leq i,j \leq n} a_{ij}^2 \Big)^{1/2},$$

Then, taking $k = 2$ above, we record the following trivial bound

$$\lambda_1^2 \leq \|A\|_F^2 . \tag{18}$$

3.2 Refined bound on spectral norms for weighted graphs

We now move on to a sharp bound on the spectral norm in terms of the Frobenius bound for networks improving the above.

Before stating the result let us discuss a situation where one already obtains an improvement over (18), namely for bipartite graphs. This is because of the underlying symmetry in the spectrum, as a consequence of which we get $\lambda_1 = -\lambda_n$ and hence

$$\lambda_1(A)^2 \leq \frac{1}{2}\|A\|_F^2 .$$

The main result of this section is a new and sharp generalization of this inequality.

Proposition 3.1 *Let k be the maximum size of clique contained in G. Then, for any conductance $a : E \to \mathbb{R}$, we have*

$$\lambda_1(A)^2 \leq \frac{k-1}{k}\|A\|_F^2 \tag{19}$$

Remark 3.2 For G a clique of size k, with adjacency matrix A, it is straightforward to see that

$$\lambda_1(A)^2 = \frac{k-1}{k}\|A\|_F^2 . \tag{20}$$

This follows from the fact that a $k \times k$ matrix whose off-diagonal entries are 1 and on-diagonal entries are 0 has the largest eigenvalue $k-1$ and the Frobenius norm $\sqrt{k^2-k}$.

The proof of the proposition will rely crucially on the following bound which goes back to the seminal work of Motzkin and Straus [37] whose proof we include for completeness.

Lemma 3.3 *Suppose that k is the maximum size of clique contained in the graph G with vertex set $[n]$. Let $f = (f_1, \cdots, f_n)$ be a vector with $\sum_{i=1}^n f_i = s$ and $f_i \geq 0$. Then,*

$$\sum_{i<j, i\sim j} f_i f_j \leq \frac{k-1}{2k} s^2. \tag{21}$$

We first provide the proof of the proposition before proving the above lemma.

Proof of Proposition 3.1 By the variational characterization of the largest eigenvalue,

$$\lambda_1(A) = \sup_{\|f\|_2=1} \sum_{i\sim j} a_{ij} f_i f_j .$$

Thus, for any conductance $a : E \to \mathbb{R}$,

$$\begin{aligned}\frac{\lambda_1(A)}{\|A\|_F} &= \sup_{\|f\|_2=1} \frac{\sum_{i\sim j} a_{ij} f_i f_j}{\|A\|_F} \\ &\le \sup_{\|f\|_2=1} \frac{(\sum_{i\sim j} a_{ij}^2)^{1/2}(\sum_{i\sim j} f_i^2 f_j^2)^{1/2}}{\|A\|_F} \\ &= \sup_{\|f\|_2=1} \Big(\sum_{i\sim j} f_i^2 f_j^2\Big)^{1/2} = \sup_{\|w\|_1=1, w_i\ge 0} \Big(\sum_{i\sim j} w_i w_j\Big)^{1/2},\end{aligned}$$

where the second line follows by Cauchy-Schwarz inequality and the final equality witnesses the $\ell_2 \to \ell_1$ reduction. By Lemma 3.3, we have

$$\sup_{\|w\|_1=1, w_i\ge 0} \Big(\sum_{i\sim j} w_i w_j\Big)^{1/2} \le \Big(\frac{k-1}{k}\Big)^{1/2},$$

which finishes the proof. □

We now provide the proof of Lemma 3.3.

Proof of Lemma 3.3 The proof is based on a 'mass transportation' argument. By homogeneity, it suffices to assume $s = 1$. We first verify (21) when G is itself a clique of size m. In other words, we claim that if $\sum_{i=1}^m f_i = 1$ and $f_i \ge 0$, then

$$\sum_{1\le i<j\le m} f_i f_j \le \frac{m-1}{2m}. \tag{22}$$

This follows from the simple equation $2\sum_{i<j} f_i f_j = (\sum_i f_i)^2 - \sum f_i^2$ and that $\sum_i f_i^2 \ge \frac{1}{m}$ (by Cauchy-Schwarz inequality).

We now prove (21) for the general graphs G. Assuming that G is not a clique of size k, one can choose two vertices v_1 and v_2 such that $v_1 \not\sim v_2$. Without loss of generality, we assume $\sum_{i\sim v_1} f_i \ge \sum_{j\sim v_2} f_j$. This allows us to transport mass from v_2 to v_1 without decreasing the objective function. Namely, since

$$\sum f_i f_j = \Big(\sum_{i\sim v_1} f_i\Big) f_{v_1} + \Big(\sum_{j\sim v_2} f_j\Big) f_{v_2} + \sum_{i,j\ne v_1,v_2, i\sim j} f_i f_j$$

is linear in f_{v_1} and f_{v_2}, f does not decrease when $f = (\cdots, f_{v_1}, \cdots, f_{v_2}, \cdots,)$ is replaced by $f^{(1)} = (\cdots, f_{v_1} + f_{v_2}, \cdots, 0, \cdots)$. After removing the zero at v_2, we obtain a new vector $\tilde{f}^{(1)}$ on the new graph G_1 obtained by deletion of the vertex v_2 and the edges incident on it.

We repeat this procedure to get a series of vectors $\tilde{f}^{(1)}, \cdots, \tilde{f}^{(\ell)}$ and graphs $G_1, \cdots, G_\ell$ such that G_{i+1} is obtained by deletion of some vertex w_{i+1} and edges incident on w_{i+1} in the graph G_i. This procedure is finished once every pair of vertices

in G_ℓ are connected, i.e. G_ℓ is a clique of size $m \le k$. This along with (22) finishes the proof. □

We end this section with a related short technical lemma which we will need later. The reader can choose to ignore this for the moment and only come back to it when it is later used.

Lemma 3.4 *Suppose that G is a tree with a vertex set $[n]$ and s, η are positive numbers. Let $v = (v_1, \cdots, v_n)$ be a vector with $\sum_i v_i = s$ and $0 \le v_i \le \eta$. Then,*

$$\sum_{i<j, i\sim j} v_i v_j \le \begin{cases} \frac{1}{4}s^2 & s < 2\eta, \\ \eta(s-\eta) & s \ge 2\eta. \end{cases} \tag{23}$$

Proof Let $\rho = \arg\max_i v_i$. Now think of the tree as rooted at ρ and orient every edge towards ρ. Thus $\sum_{i<j, i\sim j} v_i v_j \le \sum_{i \ne \rho} v_\rho v_i = v_\rho(s - v_\rho)$. Now since the function $x(s-x)$ is monotonically increasing in x for $x \le s/2$ and since $v_\rho \le \eta$, (23) follows. □

4 Upper tail large deviations: lower bound

To begin with, we state a well known estimate for the tail behavior of the maximum of Gaussian random variables which is a straightforward consequence of the following classical bound (We provide the proofs in the appendix.): For the standard Gaussian random variable X, for any $t > 0$,

$$\frac{1}{\sqrt{2\pi}} \frac{t}{t^2+1} e^{-t^2/2} \le \mathbb{P}(X > t) \le \frac{1}{\sqrt{2\pi}} \frac{1}{t} e^{-t^2/2} \tag{24}$$

(see [20, Equation (A.1)]).

Lemma 4.1 *Let $X_1, \cdots, X_m$ be i.i.d. standard Gaussian random variables and $m \ge cn$ for some constant $c > 0$. Then, there exists a constant $c' = c'(c) > 0$, such that for any $\delta > 0$,*

$$\mathbb{P}\Big(\max_{i=1,\cdots,m} X_i \ge \sqrt{2(1+\delta)\log n}\Big) \ge \frac{c'}{\sqrt{\log n}} \frac{1}{n^\delta} \tag{25}$$

and

$$\mathbb{P}\Big(\max_{i=1,\cdots,m} X_i \le \sqrt{2(1-\delta)\log n}\Big) \le e^{-c' \frac{n^\delta}{\sqrt{\log n}}}. \tag{26}$$

As indicated in Sect. 2, we first show that the number of non-zero elements of the matrix Z is at least of order n with high probability. Recall that for us $p = \frac{d}{n}$ in $\mathcal{G}_{n,p}$

throughout the article and the number of non-zero elements in X is twice the same as the number of edges in the underlying random graph G. Let us define an event

$$E_0 := \left\{ |\{1 \le i < j \le n : X_{ij} \neq 0\}| > \frac{d}{16} n \right\}. \tag{27}$$

Lemma 4.2 *There exists a constant $c > 0$ such that for sufficiently large n,*

$$\mathbb{P}(E_0^c) \le e^{-cn}.$$

This follows from standard large deviation estimates and we include the proof in the appendix for completeness.

Proof of Theorem 1.1: lower bound As indicated in Sect. 2, there is a slight distinction between $k = 2$, and $k \ge 3$, i.e. the lower bound is governed by two related but distinct events, a large value realized on an edge, or existence of a clique of size at least 3 with the Gaussians *uniformly* large on the edges in the clique.

Single large value: We first deal with the former case and prove

$$\limsup_{n\to\infty} -\frac{1}{\log n} \log \mathbb{P}(\lambda_1 \ge \sqrt{2(1+\delta)\log n}) \le \delta. \tag{28}$$

Since the matrix Z is Hermitian,

$$\lambda_1 \ge \max_{1\le i<j\le n} Z_{ij}. \tag{29}$$

Thus,

$$\begin{aligned} \mathbb{P}(\lambda_1 \ge \sqrt{2(1+\delta)\log n}) &\ge \mathbb{P}(\max_{1\le i<j\le n} Z_{ij} \ge \sqrt{2(1+\delta)\log n}) \\ &\ge \mathbb{E}\left(\mathbb{P}(\max_{1\le i<j\le n} Z_{ij} \ge \sqrt{2(1+\delta)\log n} \mid X) \mathbb{1}_{E_0} \right). \end{aligned} \tag{30}$$

By Lemma 4.1, on the event E_0,

$$\mathbb{P}(\max_{1\le i<j\le n} Z_{ij} \ge \sqrt{2(1+\delta)\log n} \mid X) \ge C \frac{1}{\sqrt{\log n}} \frac{1}{n^{\delta}}. \tag{31}$$

Thus, by (30), (31) and Lemma 4.2, we obtain (28).

Clique construction: We now move on to the clique construction. To this end, fix a positive integer m and let G be a network on the clique of size m, K_m, whose conductances $\{Y_{ij} : 1 \le i < j \le m\}$ are i.i.d. standard Gaussians. We denote by $\lambda(Y)$ the largest eigenvalue of the adjacency/conductance matrix $Y = (Y_{ij})$ of the network.

By (24), for some constant $C = C(\delta) > 0$,

$$\mathbb{P}(\lambda(Y) \ge \sqrt{2(1+\delta)\log n}) \ge \mathbb{P}\Big(Y_{ij} \ge \frac{1}{k-1}\sqrt{2(1+\delta)\log n},\ \forall 1 \le i < j \le k \Big)$$

$$\geq \left(\frac{C}{\sqrt{\log n}} n^{-\frac{1+\delta}{(k-1)^2}}\right)^{\binom{k}{2}}. \tag{32}$$

Next, we need an estimate of the probability that a graph contains a clique of size k. This is provided in the next lemma which along with (32) imply that for any $k \geq 3$,

$$\begin{aligned}\mathbb{P}(\lambda_1 \geq \sqrt{2(1+\delta)\log n}) &\geq C n^{-\binom{k}{2}+k} \left(\frac{C}{\sqrt{\log n}} n^{-\frac{1+\delta}{(k-1)^2}}\right)^{\binom{k}{2}} \\ &= C\left(\frac{C}{\sqrt{\log n}}\right)^{\binom{k}{2}} n^{-\frac{k(k-3)}{2}-\frac{1+\delta}{2}\frac{k}{k-1}} = n^{-\phi_\delta(k)+o(1)}. \end{aligned} \tag{33}$$

Since $\phi_\delta(2) = \delta$, putting (28) and (33) together, we are done. □

Lemma 4.3 *Let $k \geq 3$ be a positive integer. Then, there exists a constant $C = C(k,d) > 0$ such that the probability that $\mathcal{G}_{n,\frac{d}{n}}$ contains a clique of size k is bounded by*

$$\frac{C}{n^{\binom{k}{2}-k}}. \tag{34}$$

Proof Note that the expected number of cliques is indeed up to constants $\frac{1}{n^{\binom{k}{2}-k}}$. which implies the upper bound. Thus to lower bound the probability of existence of at least one clique we use the familiar second moment method. However as has been used several times in the probabilistic combinatorics literature (see e.g., [29, Theorem 2.3]), to control the second moment, it will be useful to work with the number of cliques which are also their respective connected components. To this end, let us denote their number by N_k. Then,

$$\begin{aligned}\mathbb{E}N_k = \binom{n}{k} p^{\binom{k}{2}} (1-p)^{k(n-k)} &\geq \frac{e^{k-1} d^{\binom{k}{2}}}{k^{k+\frac{1}{2}}} \Big(1-\frac{k}{n}\Big)^k \Big(1-\frac{d}{n}\Big)^{k(n-k)} \frac{1}{n^{\binom{k}{2}-k}} \\ &\geq C \frac{1}{n^{\binom{k}{2}-k}}. \end{aligned} \tag{35}$$

where above we use Stirling's formula to approximate $k!$ and we use the bound $n!/(n-k)! \geq (n-k)^k$. Further,

$$\mathbb{E}N_k^2 = \mathbb{E}N_k + \binom{n}{k}\binom{n-k}{k} p^{2\binom{k}{2}} (1-p)^{k^2+2k(n-2k)} \leq \mathbb{E}N_k + (1-p)^{-k^2} (\mathbb{E}N_k)^2. \tag{36}$$

Note that

$$\mathbb{E}N_k = \mathbb{E}[N_k \mathbb{1}_{N_k \geq 1}] \leq (\mathbb{E}N_k^2)^{1/2} \mathbb{P}(N_k \geq 1)^{1/2}.$$

Thus, for sufficiently large n,

$$\mathbb{P}(N_k \geq 1) \geq \frac{(\mathbb{E}N_k)^2}{\mathbb{E}N_k^2} \geq \frac{1}{(\mathbb{E}N_k)^{-1} + (1-p)^{-k^2}} \geq \frac{1}{Cn^{\binom{k}{2}-k} + 2}. \tag{37}$$

□

5 Upper tail large deviations: upper bound

A significant fraction of the novel ideas in the paper can be found in this section which aims to implement the high level strategy outlined in Sect. 2. Before beginning, we include a short roadmap to indicate what the different subsections achieve. In Sect. 5.1 we record tail estimates for sums of squares of Gaussian variables conditioned to be large. In Sect. 5.2 we show that with high probability the network $Z^{(2)}$ from Sect. 2 is spectrally negligible. We then move on to analyzing the connectivity structure of the graph $X^{(1)}$ underlying the network $Z^{(1)}$, including its maximum degree, size of its connected components and the number of tree excess edges they contain, in Sect. 5.3. In Sect. 5.4 we prove a key proposition (Proposition 5.7) establishing tails for the largest eigenvalue for tree like networks in terms of the largest clique. Finally in Sect. 5.5, we prove the upper bound in Theorem 1.1.

5.1 Chi-square tail estimates

We record the following estimate that will be crucial in our applications whose proof is provided in the appendix.

Lemma 5.1 *Let $\tilde{Y}$ be a standard Gaussian conditioned on $|\tilde{Y}| > \sqrt{\varepsilon \log\log n}$ and $\tilde{Y}_1, \cdots, \tilde{Y}_m$ be independent copies of $\tilde{Y}$. Then, there exists a universal constant $C > 0$ such that for any $L > m$ and $\varepsilon > 0$,*

$$\mathbb{P}(\tilde{Y}_1^2 + \cdots + \tilde{Y}_m^2 \geq L) \leq C^m e^{-\frac{1}{2}L} e^{\frac{1}{2}m} \Big(\frac{L}{m}\Big)^m e^{\frac{1}{2}\varepsilon m \log\log n}. \tag{38}$$

In particular, for any $a, b, c > 0$, let $m \leq b\frac{\log n}{\log\log n} + c$ and $L = a \log n$. Then, for any $\gamma > 0$, for sufficiently large n,

$$\mathbb{P}(\tilde{Y}_1^2 + \cdots + \tilde{Y}_m^2 \geq a \log n) \leq n^{-\frac{a}{2} + \frac{\varepsilon b}{2} + \gamma}. \tag{39}$$

Recall from Sect. 2, the decompositions

$$Y_{ij} = Y_{ij}^{(1)} + Y_{ij}^{(2)},$$

where $Y_{ij}^{(1)} = Y_{ij}\mathbb{1}_{|Y_{ij}|>\sqrt{\varepsilon \log\log n}}$ and similarly $Y_{ij}^{(2)} = Y_{ij}\mathbb{1}_{|Y_{ij}|\le\sqrt{\varepsilon \log\log n}}$. Thus, we can write the matrix Z as $Z^{(1)} + Z^{(2)}$ with

$$Z_{ij}^{(1)} = X_{ij}Y_{ij}^{(1)}, \quad Z_{ij}^{(2)} = X_{ij}Y_{ij}^{(2)}. \tag{40}$$

5.2 Spectrally negligible component

We next prove an upper bound on the probability that $Z^{(2)}$ has high spectral norm.

Lemma 5.2 *For $\delta > 0$,*

$$\liminf_{n\to\infty} \frac{-\log \mathbb{P}(\lambda_1(Z^{(2)}) \ge \sqrt{\varepsilon}(1+\delta)\sqrt{\log n})}{\log n} \ge 2\delta + \delta^2.$$

Proof The proof relies on the results of the previously mentioned recent work [13]. By [13, Theorem 1.1],

$$\lim_{n\to\infty} \frac{-\log \mathbb{P}(\lambda_1(X) \ge (1+\delta)\sqrt{\frac{\log n}{\log\log n}})}{\log n} = 2\delta + \delta^2.$$

Since $|Z_{ij}^{(2)}| \le X_{ij}\sqrt{\varepsilon \log\log n}$, we have $\lambda_1(Z^{(2)}) \le \sqrt{\varepsilon \log\log n} \cdot \lambda_1(X)$ which concludes the proof. □

5.3 Connectivity structure of highly sub-critical Erdos-Rényi graphs

We will now shift our focus to $Z^{(1)}$. Recall that $X_{ij}^{(1)} = X_{ij}\mathbb{1}_{|Y_{ij}|>\sqrt{\varepsilon \log\log n}}$. By the tail bound for Gaussian stated in (24), for large n, $X^{(1)}$ is distributed as $\mathcal{G}_{n,q}$ with

$$q \le \frac{d}{n}\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\varepsilon \log\log n} = \frac{d'}{n}\frac{1}{(\log n)^{\varepsilon/2}}, \tag{41}$$

where $d' = \frac{d}{\sqrt{2\pi}}$.

For any graph G, we denote by $d_1(G)$, the largest degree of G. It is proved in [32] (see also [13, Proposition 1.3]) that the typical value of $d_1(\mathcal{G}_{n,r})$ is $\frac{\log n}{\log\log n - \log(nr)}$, when

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

Furthermore, the following large deviation result is a consequence of [13, Proposition 1.3].

Lemma 5.3 *For $\delta_1 > 0$, let $\mathcal{D}_{\delta_1}$ be an event defined by*

$$\mathcal{D}_{\delta_1} := \Big\{ d_1(X^{(1)}) \le (1+\delta_1)\frac{\log n}{\log\log n} \Big\}. \tag{42}$$

Then,

$$\liminf_{n\to\infty} \frac{-\log \mathbb{P}(\mathcal{D}^c_{\delta_1})}{\log n} \ge \delta_1.$$

Proof The statement, where the inequality above is replaced with an equality, for the case $r = \frac{d}{n}$ is obtained in [13, Proposition 1.3], by plugging in $r = \frac{d}{n}$ in the latter and noting that in this case

$$\frac{\log n}{\log\log n - \log(nr)} = \frac{\log n}{\log\log n - \log d}.$$

The above result then follows by observing that $\mathcal{G}_{n,\frac{d}{n}}$ stochastically dominates $\mathcal{G}_{n,q}$ and $d_1(G)$ is an increasing function of the graph. □

We next move on to a refined analysis of the connectivity structure of the graph $X^{(1)}$. Towards this, let $C_1, \cdots, C_m$ be its connected components. The next lemma establishes a bound of the order of $\frac{\log n}{\log\log n}$ on the size of the largest component in contrast to the bounds of $\Theta(\log n)$, $\Theta(n^{2/3})$, or $\Theta(n)$, that one has for $\mathcal{G}_{n,\frac{d}{n}}$ depending on if $d < 1, d = 1$ or $d > 1$. This sub-logarithmic bound will be crucial in our application and justifies our sparsification step.

Lemma 5.4 *For $\delta_2 > 0$, let $\mathcal{C}_{\delta_2}$ be the following event.*

$$\mathcal{C}_{\delta_2} := \Big\{ |C_i| \le \frac{2+\delta_2}{\varepsilon}\frac{\log n}{\log\log n},\ \forall i \Big\}. \tag{43}$$

Then,

$$\liminf_{n\to\infty} \frac{-\log \mathbb{P}(\mathcal{C}^c_{\delta_2})}{\log n} \ge \frac{\delta_2}{2}.$$

Proof The proof implements the standard first moment argument, (see e.g., [16, Chapter 5,6]). Throughout the proof, the value of the constant C may change from line to line. Let $\bar{N}_{k,-1}$ be the number of *connected* subgraphs having k vertices and $k-1$ edges, in other words the number of trees of size k. Using (41) and Stirling's formula, and the fact that the number of labelled spanning trees on k vertices is k^{k+2}, for some large constant $c_0 > 0$,

$$\mathbb{E}\bar{N}_{k,-1} \le \binom{n}{k} k^{k+2} \Big(\frac{d'}{n} \frac{1}{(\log n)^{\varepsilon/2}} \Big)^{k-1}$$

$$\leq Ce^k \frac{n^k}{k^k} k^{k+2} \frac{(d')^{k-1}}{n^{k-1}(\log n)^{\frac{\varepsilon}{2}(k-1)}} = Cn \frac{e^k k^2 (d')^{k-1}}{(\log n)^{\frac{\varepsilon}{2}(k-1)}}$$
$$\leq Cn(\log n)^{\varepsilon/2} \Big(\frac{c_0}{(\log n)^{\varepsilon/2}} \Big)^k. \tag{44}$$

Hence, denoting N_k by the number of connected components with k vertices, picking a spanning tree from each connected component, one obtains

$$\mathbb{E}N_k \leq \mathbb{E}\bar{N}_{k,-1} \leq Cn(\log n)^{\varepsilon/2} \Big(\frac{c_0}{(\log n)^{\varepsilon/2}} \Big)^k.$$

Define $m := \frac{2+\delta_2}{\varepsilon} \frac{\log n}{\log\log n}$, and let N be the number of connected components having at least m vertices. Then,

$$\mathbb{E}N = \mathbb{E}\sum_{k=m}^{n} N_k \leq Cn(\log n)^{\varepsilon/2} \Big(\frac{c_0}{(\log n)^{\varepsilon/2}} \Big)^m \leq C(\log n)^{\varepsilon/2} n^{\frac{(\log c_0)(2+\delta_2)}{\varepsilon \log\log n}} n^{-\frac{\delta_2}{2}}.$$

Since $\mathbb{P}(N \geq 1) \leq \mathbb{E}(N)$, the proof is complete. □

For our applications, we will also need to bound the number of subgraphs having k vertices and $k + \ell$ edges without the subgraph necessarily being connected. This estimate will be crucially used later to prove the structure theorem conditioned on $\mathcal{U}_\delta$.

Lemma 5.5 *For $\ell \geq 0$, let $N_{k,\ell}$ be the number of subgraphs in $X^{(1)}$ having k vertices and $k + \ell$ edges. Then, for $0 \leq \ell \leq \binom{k}{2} - k$,*

$$\mathbb{E}N_{k,\ell} \leq C \min\left(\Big(\frac{k}{n}\Big)^\ell, \Big(\frac{d'e^2}{(\log n)^{\varepsilon/2}}\Big)^{k+\ell} \right).$$

Proof Denote by $C_{k,\ell}$ the number of labelled graphs with k vertices and $k + \ell$ edges. Then, for any $-k \leq \ell \leq \binom{k}{2} - k$, using Stirling's formula we have

$$C_{k,\ell} = \binom{\binom{k}{2}}{k+\ell} \leq \binom{k^2}{k+\ell} \leq \frac{(k^2)^{k+\ell}}{(k+\ell)!} \leq e^{k+\ell} \frac{k^{2(k+\ell)}}{(k+\ell)^{k+\ell}}. \tag{45}$$

Then, for $0 \leq \ell \leq \binom{k}{2} - k$,

$$\mathbb{E}N_{k,\ell} \leq \binom{n}{k} C_{k,\ell} q^{k+\ell} \overset{(41)}{\leq} Ce^{2k+\ell} \frac{n^k}{k^k} \frac{k^{2(k+\ell)}}{(k+\ell)^{k+\ell}} \Big(\frac{d'}{n} \frac{1}{(\log n)^{\varepsilon/2}} \Big)^{k+\ell}$$
$$\leq C\Big(\frac{k}{n}\Big)^\ell \Big(\frac{d'e^2}{(\log n)^{\varepsilon/2}} \Big)^{k+\ell}. \tag{46}$$

where in the first inequality we use Stirling's formula again to bound $k!$. In particular, since $k \le n$,

$$\mathbb{E}N_{k,\ell} \le C\Big(\frac{d'e^2}{(\log n)^{\varepsilon/2}}\Big)^{k+\ell}, \tag{47}$$

and since $d'e^2 \le (\log n)^{\varepsilon/2}$ for sufficiently large n,

$$\mathbb{E}N_{k,\ell} \le C\Big(\frac{k}{n}\Big)^{\ell}. \tag{48}$$

□

Having bounded the maximum component size, we next proceed to estimating how close the components are to trees by bounding the number of tree excess edges, i.e., how many edges need to be removed from such a component to obtain a tree.

Lemma 5.6 *For $\delta_3 > 0$, let $\mathcal{E}_{\delta_3}$ be the event defined by*

$$\mathcal{E}_{\delta_3} := \{|E(C_i)| < |V(C_i)| + \delta_3,\ \forall i\}. \tag{49}$$

Then,

$$\liminf_{n\to\infty} \frac{-\log \mathbb{P}(\mathcal{E}_{\delta_3}^c)}{\log n} \ge \delta_3. \tag{50}$$

In addition, define the event $\mathcal{T}$ by

$$\mathcal{T} := \{|E(C_i)| = |V(C_i)| - 1,\ \forall i\}.$$

In other words, $\mathcal{T}$ is the event that all the connected components of $X^{(1)}$ are trees. Then,

$$\mathbb{P}(\mathcal{T}^c) \le \frac{C}{(\log n)^{\varepsilon}}. \tag{51}$$

Proof For $\ell \ge 0$, recall the notation $N_{k,\ell}$ from Lemma 5.5. Since the occurrence of the event $\mathcal{E}_{\delta_3}^c \cap \mathcal{C}_{2\delta_3}$ demands the existence of a connected component C_i with $|C_i| \le \Big\lfloor \frac{2+2\delta_3}{\varepsilon} \frac{\log n}{\log\log n} \Big\rfloor =: m$ and $|E(C_i)| \ge |C_i| + \lceil \delta_3 \rceil$, by the first moment bound,

$$\mathbb{P}(\mathcal{E}_{\delta_3}^c \cap \mathcal{C}_{2\delta_3}) \le \sum_{k=3}^{m} \sum_{\ell=\lceil\delta_3\rceil}^{\binom{k}{2}-k} \mathbb{E}N_{k,\ell} \overset{(48)}{\le} C \sum_{k=3}^{m} \Big(\frac{k}{n}\Big)^{\lceil\delta_3\rceil} \le C\frac{m^{\lceil\delta_3\rceil+1}}{n^{\lceil\delta_3\rceil}}. \tag{52}$$

Therefore, by (52) and Lemma 5.4 (with $\delta_2 = 2\delta_3$), we obtain (50).

Next, we prove (51). Let N_{cycle} be the number of cycles in $X^{(1)}$. Then,

$$\mathbb{E}N_{\text{cycle}} = \sum_{k=3}^{n} \binom{n}{k} \frac{(k-1)!}{2} q^k \le \sum_{k=3}^{n} \frac{n^k}{2k} \Big(\frac{d'}{n} \frac{1}{(\log n)^{\varepsilon/2}} \Big)^k \le \frac{C}{(\log n)^{\varepsilon}}.$$

Since the occurrence of $\mathcal{T}^c$ implies the existence of cycle, by the first moment bound, we obtain (51). □

5.4 Spectral tail for tree like networks

We have so far defined the events $\mathcal{D}_\alpha$, $\mathcal{C}_\alpha$, $\mathcal{E}_\alpha$, $\mathcal{T}$, and in the previous series of lemmas, having established that each connected component is of size $O(\frac{\log n}{\log\log n})$ and the number of excess edges is bounded with high probability, in the following key proposition, we control the spectral norm of such a connected component. This will be a particularly important ingredient in the proof of Theorem 1.1.

Proposition 5.7 *Consider a connected network $G = (V, E, A)$ (where $A = (a_{ij})$ is the matrix of conductances) satisfying the following properties:*

1. $d_1(G) \le c_1 \frac{\log n}{\log\log n}$
2. $|V| \le c_2 \frac{\log n}{\log\log n}$
3. $|E| \le |V| + c_3$

Suppose that the conductance matrix A is given by i.i.d. Gaussians associated to each element of E, conditioned on having absolute value greater than $\sqrt{\varepsilon \log\log n}$. Let k be a maximum size of clique in G and λ be the largest eigenvalue of A. Then, for any $\varepsilon, \alpha, \gamma, \eta > 0$ with $\eta < \frac{1}{2}$, for sufficiently large n,

$$\mathbb{P}(\lambda \ge \sqrt{2\alpha \log n}) \le n^{-\frac{\alpha}{2\theta^2} + \frac{\varepsilon c_2}{2} + \gamma} + n^{-\frac{k}{2(k-1)}(1-\theta)^2\alpha + \frac{c_1\varepsilon}{2\eta^2} + \gamma}, \tag{53}$$

where $\theta := (2\eta^2 + 2\eta^4 c_3)^{1/4}$.

The expression on the right hand side is technical but the constants $\varepsilon, \eta, \gamma$ will be suitably chosen sufficiently close to zero so that $n^{-\frac{\alpha}{2\theta^2} + \frac{\varepsilon c_2}{2} + \gamma}$ and $n^{\frac{c_1\varepsilon}{2\eta^2} + \gamma}$ are negligible and the dominant behavior will be $n^{-\frac{k}{2(k-1)}\alpha}$.

From now on, for any graph H, we denote by $E(H)$ and $\overrightarrow{E(H)}$ the sets of undirected and directed edges in H respectively. Recall that E denotes the set of edges in G, and let $\overrightarrow{E}$ be its directed version.

Proof The proof proceeds by analyzing the leading eigenvector (A priori, there may be several such eigenvectors and in this case, we choose any one of them arbitrarily. However, in fact, owing to the continuity of the Gaussian distribution, one can show that almost surely, every non-zero eigenvalue is simple. See Remark 5.8 for further elaboration). Let $V = [\ell]$ and $f = (f_1, \cdots, f_\ell)$ be the unit (random) eigenvector associated with the largest eigenvalue $\lambda := \lambda_1(G)$. Thus by definition, $\lambda = f^\top A f$.

One would have liked to use Proposition 3.1 and the tail estimate (39). However the application of the latter is useful only when the parameter b in the upper bound of m is small enough compared to $\frac{1}{\varepsilon}$. On the other hand, in Lemma 5.4, the bound on $|C_i|$ which would be m in the application is $O(\frac{1}{\varepsilon}\frac{\log n}{\log\log n})$ rendering the above straightforward strategy useless. To address this, the first step is to argue that entries of f that are small in absolute value do not contribute much to the above quadratic form. This allows us to focus on only the large entries, of which there are not too many and hence allows an application of the above outlined strategy with a reduced value of m. Towards this, for $0 < \eta < 1/2$, define the collection of vertices

$$I := \{i \in [\ell] : f_i^2 < \eta^2\}.$$

Let B_1 be the collection of (directed) edges defined by

$$B_1 := \{(i, j) \in \overrightarrow{E} : i, j \in I\}$$

and let $B_2 := \overrightarrow{E} \setminus B_1$ where again each edge is considered twice (this is done simply as a matter of convention) Now since f is a unit vector, by Markov's inequality, $|I^c| \le \frac{1}{\eta^2}$. In addition, by the upper bound on the max-degree in condition (1), we obtain

$$\frac{1}{2}|B_2| \le \frac{c_1}{\eta^2}\frac{\log n}{\log\log n}. \tag{54}$$

We write

$$\lambda = \sum_{(i,j)\in\overrightarrow{E}} a_{ij} f_i f_j = \sum_{(i,j)\in B_1} a_{ij} f_i f_j + \sum_{(i,j)\in B_2} a_{ij} f_i f_j =: S_1 + S_2.$$

Recall $\theta = (2\eta^2 + 2\eta^4 c_3)^{1/4}$, we have

$$\mathbb{P}(\lambda \ge \sqrt{2\alpha \log n}) \le \mathbb{P}(S_1 \ge \theta\sqrt{2\alpha \log n}) + \mathbb{P}(S_2 \ge (1-\theta)\sqrt{2\alpha \log n}). \tag{55}$$

Of course, the above inequality holds for any θ and the particular choice we make is guided by our subsequent estimates of S_1 and S_2. First, we show that

$$\sum_{(i,j)\in B_1} f_i^2 f_j^2 \le 2\eta^2 + 2\eta^4 c_3 = \theta^4. \tag{56}$$

We will rely on Lemma 3.4. Choose a spanning tree T of G, and define a set of (directed) edges $\overrightarrow{E'} := \overrightarrow{E} \setminus \overrightarrow{E(T)}$. Then, by condition (3) on the number of excess edges, $\frac{1}{2}|E'| \le c_3 + 1$. Now the graph with edge set $B_1 \setminus \overrightarrow{E'}$ is necessarily a forest. Since adding more edges can only increase $\sum_{(i,j)\in B_1\setminus\overrightarrow{E'}} f_i^2 f_j^2$ we can in fact assume

that the graph with edge set $B_1 \setminus \overrightarrow{E'}$ is a tree. Now applying Lemma 3.4 with $s = 1$, and since $2\eta^2 \le 1$, we conclude that

$$\sum_{(i,j)\in B_1\setminus\overrightarrow{E'}} f_i^2 f_j^2 \le 2\eta^2(1-\eta^2). \tag{57}$$

Hence,

$$\sum_{(i,j)\in B_1} f_i^2 f_j^2 = \sum_{(i,j)\in B_1\setminus\overrightarrow{E'}} f_i^2 f_j^2 + \sum_{(i,j)\in B_1\cap\overrightarrow{E'}} f_i^2 f_j^2 \le 2\eta^2(1-\eta^2) + 2(c_3+1)\eta^4, \tag{58}$$

where for the second term, we simply use the fact that the total number of summands is at most $2(c_3+1)$ with each being at most η^4. This proves (56). Hence by the definition of S_1, by Cauchy–Schwarz inequality, we immediately have

$$S_1 \le \Big(\sum_{(i,j)\in B_1} a_{ij}^2\Big)^{1/2}\Big(\sum_{(i,j)\in B_1} f_i^2 f_j^2\Big)^{1/2} \le \theta^2\Big(\sum_{(i,j)\in B_1} a_{ij}^2\Big)^{1/2} \le \theta^2\Big(\sum_{(i,j)\in E} a_{ij}^2\Big)^{1/2}.$$

Thus, for any $\gamma > 0$, for sufficiently large n,

$$\mathbb{P}(S_1 \ge \theta\sqrt{2\alpha\log n}) \le \mathbb{P}\Big(\sum_{i<j,(i,j)\in E} a_{ij}^2 \ge \frac{\alpha}{\theta^2}\log n\Big) \le n^{-\frac{\alpha}{2\theta^2}+\frac{\varepsilon c_2}{2}+\gamma}, \tag{59}$$

where the last inequality follows by a direct application of (39) in Lemma 5.1, with $L = \frac{\alpha}{\theta^2}\log n$ and $m = c_2\frac{\log n}{\log\log n} + c_3$.

Next, we estimate S_2. Since, by hypothesis, the maximum size of clique in the subgraph induced by edges in B_2 is no larger than k, by Lemma 3.3,

$$S_2 \le \Big(\sum_{(i,j)\in B_2} a_{ij}^2\Big)^{1/2}\Big(\sum_{(i,j)\in B_2} f_i^2 f_j^2\Big)^{1/2} \le \Big(\frac{k-1}{k}\Big)^{1/2}\Big(\sum_{(i,j)\in B_2} a_{ij}^2\Big)^{1/2}. \tag{60}$$

Note that the event $\sum_{i<j(i,j)\in B_2} a_{ij}^2 \ge t$ implies the existence of a random subset $J \in [n]$ with $|J| \le \left\lfloor \frac{1}{\eta^2} \right\rfloor$ such that $\sum_{i \text{ or } j\in J, i<j, i\sim j} a_{ij}^2 \ge t$. Hence, for any $\gamma > 0$, for sufficiently large n,

$$\begin{aligned}\mathbb{P}(S_2 \ge (1-\theta)\sqrt{2\alpha\log n}) &\overset{(60)}{\le} \mathbb{P}\Big(\sum_{i<j,\,(i,j)\in B_2} a_{ij}^2 \ge \frac{k}{k-1}(1-\theta)^2\alpha\log n\Big)\\ &\le |V|^{\left\lfloor\frac{1}{\eta^2}\right\rfloor} n^{-\frac{k}{2(k-1)}(1-\theta)^2\alpha+\frac{c_1\varepsilon}{2\eta^2}+\frac{\gamma}{2}} \le n^{-\frac{k}{2(k-1)}(1-\theta)^2\alpha+\frac{c_1\varepsilon}{2\eta^2}+\gamma}.\end{aligned} \tag{61}$$

The second inequality is obtained by a simple first moment bound, in conjunction with (39) in Lemma 5.1 with $L = \frac{k}{k-1}(1-\theta)^2\alpha \log n$ and $m \le \frac{c_1}{\eta^2}\frac{\log n}{\log\log n}$ (see (54)). In the last inequality, we used condition (2) i.e., $|V| \le c_2 \frac{\log n}{\log\log n}$, to bound the term $|V|^{\left\lfloor \frac{1}{\eta^2} \right\rfloor}$ by $n^{\gamma/2}$ for sufficiently large n.

Thus, by (55), (59) and (61), for sufficiently large n,

$$\mathbb{P}(\lambda \ge \sqrt{2\alpha \log n}) \le n^{-\frac{\alpha}{2\theta^2}+\frac{\varepsilon c_2}{2}+\gamma} + n^{-\frac{k}{2(k-1)}(1-\theta)^2\alpha+\frac{c_1\varepsilon}{2\eta^2}+\gamma} \tag{62}$$

which finishes the proof. □

Remark 5.8 (Simplicity of the non-zero spectrum). One can prove that the conductance matrix (denoted by A) of any network with i.i.d. Gaussian edge weights on a graph $G = (V, E)$ has no *non-zero* multiple eigenvalues almost surely. Since weights are continuous random variables, it is natural to expect that, almost surely, every eigenvalue is simple. However, one needs to be careful since networks having several isolated vertices possess multiple zero-eigenvalues and hence what turns out to be true is the simplicity of every *non-zero* eigenvalue. We provide a brief outline below on how to formalize this.

As an intermediate step, using induction on the number of edges, we first show that for any fixed $\lambda \neq 0$, almost surely

$$\det(A - \lambda I) \neq 0 \tag{63}$$

(it is obvious that the base case, all graphs without any edges, satisfies this property). Take any edge $e = (v, w)$ and let X be the Gaussian weight on it. Conditioned on the weights on all the other edges, $\det(A - \lambda I)$ is (at most) a quadratic function in X. Denoting by $\bar{A}$ a conductance matrix induced by $V\backslash\{v, w\}$, the coefficient of X^2 in $\det(A-\lambda I)$ is given by $-\det(\bar{A}-\lambda I)$, which is non-zero by the induction hypothesis. Since for any $a, b, c \in \mathbb{R}$ with $a \neq 0$, $\mathbb{P}(aX^2 + bX + c = 0) = 0$, we obtain (63).

Equipped with (63), using again an induction on the number of edges, we now prove that every non-zero eigenvalue of A is simple (the base case, a graph without any edges, has no non-zero eigenvalue). Decompose the underlying graph G into connected components. By (63) and the independence of weights, non-zero eigenvalues coming from different components are distinct almost surely. Hence, without loss of generality, one can assume that G is connected.

For a vertex $v \in V$, let A_v be the conductance matrix on the network induced by $V\backslash\{v\}$. Suppose that A has a non-simple eigenvalue λ. Then, by the interlacing property of eigenvalues, for any $v \in V$, λ is also an eigenvalue of A_v.

To carry out the argument, we will rely on the claim that there exists $x \in V$ with $\lambda-$eigenvector $f^{(x)}$ of A_x which does not all vanish on the neighbors of x. Taking any $v \in V$, if $\lambda-$eigenvector $f^{(v)}$ of A_v vanishes on the neighbors of v, then take vertices $x, y \neq v$ with $x \sim y$ such that $f^{(v)}(x) = 0$ and $f^{(v)}(y) \neq 0$ (this is possible since the underlying graph is connected). One can check that the vector $f^{(x)}$ defined by $f^{(x)}(z) := f^{(v)}(z)$ for $z \in V\backslash\{x, v\}$ and $f^{(x)}(z) := 0$ for $z = v$ is an eigenvector of

A_x corresponding to the *same* eigenvalue λ. Since $f^{(x)}(y) = f^{(v)}(y) \neq 0$ and $x \sim y$, $f^{(x)}$ satisfies the desired property.

Now, for each λ-eigenvector g of A, and any x such as above, write $g = (\bar{g}_x, g_x)$, where $\bar{g}_x$ and g_x denote the $V\backslash\{x\}$ and x-coordinates of g respectively. We consider the following two possible cases and analyze each case.

1. $g_x = 0$ for all λ-eigenvectors g of A.
2. There exists a λ-eigenvector g of A such that $g_x \neq 0$.

First case: We use the simple fact that if a λ-eigenvector g of A satisfies $g_x = 0$, then $\bar{g}_x$ is a λ-eigenvector of A_x. This implies that in the first case, multiplicity property of the eigenvalue λ of the network A is passed on to the (strictly) smaller network A_x. Hence, we are done by the induction hypothesis.

Second case: Write

$$A = \begin{pmatrix} A_x & W_x \\ W_x^T & 0 \end{pmatrix},$$

where W_x denotes the column vector induced by weights on the edges having x as an endpoint. Using the eigenvalue equation, it is not hard to see that if there is a $\lambda-$eigenvector g of A with $g_x \neq 0$, then any $\lambda-$eigenvector of A_x is orthogonal to W_x. Recall that there exists a $\lambda-$eigenvector $f^{(x)}$ of A_x which does not vanish on all the neighbors of x. Since edge weights are continuous random variables and independent, for such $f^{(x)}$, $\mathbb{P}(f^{(x)} \cdot W_x = 0) = 0$ (by conditioning on the weights in A_x). Since x is random, a simple union bound over all possible choices of x concludes the proof.

With all this preparation, we are now ready to prove the upper bound in Theorem 1.1.

5.5 Proof of Theorem 1.1: upper bound

Recall the matrices from (40) as well as the matrix $X^{(1)}$ from (41). Let $C_1, \cdots, C_m$ be the connected components of $X^{(1)}$, and define $\lambda_1(C_i)$ to be the largest eigenvalue of the matrix $Z^{(1)}$ restricted to C_i. Let $Few - cycles$ be the event defined by

$$Few - cycles := \{|\{i : C_i \text{ not tree}\}| < \log n\}.$$

By Lemma 5.6 (51), the probability of existence of some cycle is $\frac{C}{(\log n)^{\varepsilon}}$. Since the occurrence of the event $Few - cycles^c$ demands the disjoint occurrence of $\log n$ many cycles, by the above fact and Van-den Berg-Kesten (BK) inequality [38],

$$\mathbb{P}(Few - cycles^c) \leq \frac{C}{(\log n)^{\varepsilon \log n/2}}. \tag{64}$$

Also, since $\lambda_1(Z) \leq \lambda_1(Z^{(1)}) + \lambda_1(Z^{(2)})$,

$$\mathbb{P}(\lambda_1(Z) \geq \sqrt{2(1+\delta)\log n})$$

$$\leq \mathbb{P}(\lambda_1(Z^{(1)}) \geq \sqrt{2(1+\delta')\log n}) + \mathbb{P}(\lambda_1(Z^{(2)}) \geq \sqrt{\varepsilon}(1+\delta)\sqrt{\log n}), \tag{65}$$

where $\delta' > 0$ is defined by

$$\sqrt{2(1+\delta')} = \sqrt{2(1+\delta)} - \sqrt{\varepsilon}(1+\delta). \tag{66}$$

Note that from this (by rearranging and multiplying both sides by $\sqrt{2(1+\delta)} + \sqrt{2(1+\delta')}$), we have

$$\delta - \sqrt{2\varepsilon}(1+\delta)^{3/2} \leq \delta' \leq \delta. \tag{67}$$

Using the result in Sect. 5.2, the second term in (65) will be negligible, so we focus on estimating the first one. Recalling $X_{ij}^{(1)} := X_{ij}\mathbb{1}_{|Y_{ij}|>\sqrt{\varepsilon \log\log n}}$, let us estimate the conditional probability $\mathbb{P}(\lambda_1(Z^{(1)}) \geq \sqrt{2(1+\delta')\log n}|X^{(1)})$ on the high probability event $\mathcal{D}_{4\delta'} \cap \mathcal{C}_{4\delta'} \cap \mathcal{E}_{4\delta'} \cap Few-cycles$. By definition, on this event, we have

$$d_1(X^{(1)}) < (1+4\delta')\frac{\log n}{\log\log n}, \tag{68}$$

$$|V(C_i)| < \frac{2+4\delta'}{\varepsilon}\frac{\log n}{\log\log n}, \quad i = 1, \cdots, m, \tag{69}$$

$$|E(C_i)| < |V(C_i)| + 4\delta', \quad i = 1, \cdots, m, \text{ and,} \tag{70}$$

$$|\{i = 1, \cdots, m : C_i \text{ not tree}\}| < \log n. \tag{71}$$

From now on we will denote by $Z_i^{(1)}$, the matrix $Z^{(1)}$ restricted to C_i, and by k_i the size of the largest clique in C_i. By (68)–(70) and Proposition 5.7 with

$$c_1 = 1+4\delta',\ c_2 = \frac{2+4\delta'}{\varepsilon},\ c_3 = 4\delta',\ \alpha = 1+\delta',\ \eta = \varepsilon^{1/4} \text{ and } \gamma = \varepsilon,$$

setting $\xi := (2\varepsilon^{1/2} + 8\varepsilon\delta')^{1/4}$, on the event $\mathcal{D}_{4\delta'} \cap \mathcal{C}_{4\delta'} \cap \mathcal{E}_{4\delta'}$, sufficiently small $\varepsilon > 0$,

$$\mathbb{P}(\lambda_1(Z_i^{(1)}) \geq \sqrt{2(1+\delta')\log n} \mid X^{(1)}) < Cn^{-\frac{k_i}{2(k_i-1)}(1-\xi)^2(1+\delta') + \frac{1+4\delta'}{2}\varepsilon^{1/2}+\varepsilon}, \tag{72}$$

by observing that for ε small enough, the first term in (53) is negligible compared to the second term and can be absorbed in the constant C. More precisely, using the bound (67), one can take sufficiently small ε such that

$$1+\delta' > 2(2\varepsilon^{1/2} + 8\varepsilon\delta')^{1/2}(1+\delta' + (1+2\delta')). \tag{73}$$

Then, for $k \geq 2$, $\frac{1+\delta'}{2\xi^2} - (1+2\delta') \geq 1+\delta' \geq \frac{k}{2(k-1)}(1-\xi)^2(1+\delta')$, which implies that the first term in (53) decays faster than the second term.

Define

$$I := \{i = 1, \cdots, m : k_i \geq 3\}, \quad J := \{i = 1, \cdots, m : k_i = 2\}, \text{ and,}$$
$$\bar{k} := \max\{k_1, \cdots, k_m\}. \tag{74}$$

Then, since $\frac{k}{k-1}$ is decreasing in k, by (72), under the event $\mathcal{D}_{4\delta'} \cap \mathcal{C}_{4\delta'} \cap \mathcal{E}_{4\delta'}$, for any $i \in I$,

$$\mathbb{P}(\lambda_1(Z_i^{(1)}) \geq \sqrt{2(1+\delta')\log n} \mid X^{(1)}) < Cn^{-\frac{\bar{k}}{2(\bar{k}-1)}(1-\xi)^2(1+\delta')+\frac{1+4\delta'}{2}\varepsilon^{1/2}+\varepsilon}, \tag{75}$$

and for any $i \in J$,

$$\mathbb{P}(\lambda_1(Z_i^{(1)}) \geq \sqrt{2(1+\delta')\log n} \mid X^{(1)}) < Cn^{-(1-\xi)^2(1+\delta')+\frac{1+4\delta'}{2}\varepsilon^{1/2}+\varepsilon}. \tag{76}$$

Also, by Lemmas 5.3, 5.4, 5.6 and (64), defining the event

$$\mathcal{F}_0 := \mathcal{D}_{4\delta'} \cap \mathcal{C}_{4\delta'} \cap \mathcal{E}_{4\delta'} \cap Few - cycles, \tag{77}$$

we have

$$\mathbb{P}(\mathcal{F}_0^c) \leq \frac{C}{n^{2\delta'}}. \tag{78}$$

Using (41), by the first moment bound, for $k \geq 3$,

$$\mathbb{P}(X^{(1)} \text{ contains a clique of size } k) \leq \binom{n}{k} q^{\binom{k}{2}} \leq \frac{(d')^{\binom{k}{2}}}{n^{\binom{k}{2}-k}}. \tag{79}$$

Also, since any connected component C_i which is a tree has $k_i = 2$, on the event $Few - cycles$, we have $|I| < \log n$. Thus, using (79) and the fact $\lambda_1(Z^{(1)}) = \max_{i=1,\cdots,m} \lambda_1(Z_i^{(1)})$,

$$\begin{aligned}
&\mathbb{P}(\lambda_1(Z^{(1)}) \geq \sqrt{2(1+\delta')\log n}) \\
&\leq \sum_{k=3}^{n} \mathbb{E}\left[\mathbb{P}(\max_{i\in I}\{\lambda_1(Z_i^{(1)})\} \geq \sqrt{2(1+\delta')\log n} \mid X^{(1)}) \mathbb{1}_{\mathcal{F}_0} \mathbb{1}_{\bar{k}=k}\right] \\
&+ \mathbb{E}\left[\mathbb{P}(\max_{i\in J}\{\lambda_1(Z_i^{(1)})\} \geq \sqrt{2(1+\delta')\log n} \mid X^{(1)}) \mathbb{1}_{\mathcal{D}_{4\delta'}\cap\mathcal{C}_{4\delta'}\cap\mathcal{E}_{4\delta'}}\right] + \mathbb{P}(\mathcal{F}_0{}^c) \\
&\leq C\log n \sum_{k=3}^{n} (d')^{\binom{k}{2}} n^{-\binom{k}{2}+k-\frac{k}{2(k-1)}(1-\xi)^2(1+\delta')+\frac{1+4\delta'}{2}\varepsilon^{1/2}+\varepsilon} \\
&+ Cn \cdot n^{-(1-\xi)^2(1+\delta')+\frac{1+4\delta'}{2}\varepsilon^{1/2}+\varepsilon} + Cn^{-2\delta'},
\end{aligned} \tag{80}$$

where (75) and (76) are used to bound the first and second terms respectively. The multiplicative factors of $\log n$ and n appear as a result of a union bound over the components contributing to the index sets I and J respectively. Recalling $\xi = (2\varepsilon^{1/2} + 8\varepsilon\delta')^{1/4}$ and δ' from (66), note that $\lim_{\varepsilon\to 0}\delta' = \delta$ and $\lim_{\varepsilon\to 0}\xi = 0$. Furthermore, recall from (3) that $\psi(\delta) = \min_{k\geq 2}\phi_k(\delta)$ where $\phi_k(\delta) = \frac{k(k-3)}{2} + \frac{1+\delta}{2}\frac{k}{k-1}$.

Hence, by bounding the term $\log n$ by n^{ε}, there exists $\eta_1 = \eta_1(\varepsilon)$ with $\lim_{\varepsilon\to 0}\eta_1 = 0$ such that the first term of RHS in (80) is bounded by

$$\sum_{k=3}^{n}(d')^{\binom{k}{2}}n^{-\binom{k}{2}+k-\frac{k}{2(k-1)}(1-\xi)^2(1+\delta')+\frac{1+4\delta'}{2}\varepsilon^{1/2}+2\varepsilon} \tag{81}$$

$$\leq \sum_{k=3}^{n}(d')^{\binom{k}{2}}n^{-\frac{k(k-3)}{2}-\frac{1+\delta}{2}\frac{k}{k-1}+\frac{\eta_1}{2}} \tag{82}$$

$$\leq C(\log n)^{1/4}(d')^{\binom{(\log n)^{1/4}}{2}}n^{-\psi(\delta)+\frac{\eta_1}{2}} + \sum_{k=(\log n)^{1/4}}^{n} n^{\frac{1}{2}\binom{k}{2}-\frac{k(k-3)}{2}-\frac{1+\delta}{2}\frac{k}{k-1}+\frac{\eta_1}{2}}$$

$$< Cn^{-\psi(\delta)+\eta_1}. \tag{83}$$

As the reader perhaps already notices, the cutoff $(\log n)^{1/4}$ is not special and any poly-log cutoff $(\log n)^r$ with $0 < r < 1/2$ works.

For further applications later, we provide a quantitative bound for η_1. Using (67) and the fact that $\frac{k}{2(k-1)} \leq 1$ for $k \geq 2$, one can estimate the difference between the two exponents of n in (81) and (82):

$$\begin{aligned}
&\frac{k}{2(k-1)}(1+\delta) - \frac{k}{2(k-1)}(1-\xi)^2(1+\delta') + \frac{1+4\delta'}{2}\varepsilon^{1/2} + 2\varepsilon \\
&\leq (1+\delta) - (1-2\xi)(1+\delta-\sqrt{2\varepsilon}(1+\delta)^{3/2}) + \frac{1+4\delta}{2}\varepsilon^{1/2} + 2\varepsilon \\
&\leq 4(\varepsilon^{1/8}+\delta^{1/4}\varepsilon^{1/4})(1+\delta) + \sqrt{2\varepsilon}(1+\delta)^{3/2} + \frac{1+4\delta}{2}\varepsilon^{1/2} + 2\varepsilon =: r_\delta(\varepsilon),
\end{aligned} \tag{84}$$

where we used $\xi = (2\varepsilon^{1/2} + 8\varepsilon\delta')^{1/4} \leq 2\varepsilon^{1/8} + 2\delta^{1/4}\varepsilon^{1/4}$ in the last inequality. In addition, for any constant $\eta_1 > 0$, the inequality (83) holds for sufficiently large n. Hence, $\eta_1 > 0$ can be chosen as

$$\eta_1 = 2r_\delta(\varepsilon), \tag{85}$$

which obviously converges to 0 as $\varepsilon \to 0$.

Similarly, for some $\eta_2 = \eta_2(\varepsilon)$ such that $\lim_{\varepsilon\to 0}\eta_2 = 0$,

$$n \cdot n^{-(1-\xi)^2(1+\delta')+\frac{1+4\delta'}{2}\varepsilon^{1/2}+\varepsilon} \leq n^{-\delta+\eta_2} \leq n^{-\psi(\delta)+\eta_2}. \tag{86}$$

Hence, applying (83) and (86) to (80), using the bound for δ' in (67), for sufficiently small ε,

$$\mathbb{P}(\lambda_1(Z^{(1)}) \geq \sqrt{2(1+\delta')\log n}) < Cn^{-\psi(\delta)+\max(\eta_1,\eta_2)}. \tag{87}$$

Recall by Lemma 5.2, for all large n,

$$\mathbb{P}(\lambda_1(Z^{(2)}) \geq \sqrt{\varepsilon}(1+\delta)\sqrt{\log n}) \leq n^{-2\delta-\delta^2+o(1)} \leq n^{-\delta+o(1)} \leq n^{-\psi(\delta)+o(1)}.$$

Since $\varepsilon > 0$ is arbitrary small, by (65) and the above two displays, we are done. □

6 Structure conditioned on $\mathcal{U}_\delta$

We prove Theorem 1.4 in this section. We begin by stating some facts about ϕ_δ. Recall that $\mathcal{M}(\delta)$ is the set of of minimizers of $\phi_\delta(\cdot)$, and by the strict convexity of $\phi_\delta(\cdot)$, $\mathcal{M}(\delta)$ is at most of size 2 containing either a single element or two consecutive numbers. In addition, since $\delta > \delta_2$, we have $\psi(\delta) < \phi_\delta(2) = \delta$. From this, one can deduce that there exists a constant $c(\delta) \in (0, \min(\delta - \psi(\delta), 1))$ such that

$$k \notin \mathcal{M}(\delta) \Rightarrow \phi_\delta(k) - \psi(\delta) \geq c(\delta) \tag{88}$$

(recall that $\psi(\delta) = \min_{k\geq 2} \phi_\delta(k)$). In fact, let us define, in the case when $\mathcal{M}(\delta) = \{h(\delta)\}$ is a singleton, by the strict convexity of $\phi_\delta(\cdot)$,

$$c(\delta) = \min\Big(\phi_\delta(h(\delta)-1) - \phi_\delta(h(\delta)), \phi_\delta(h(\delta)+1) - \phi_\delta(h(\delta)), \frac{1}{2}(\delta - \psi(\delta)), \frac{1}{2}\Big),$$

and when $\mathcal{M}(\delta) = \{h(\delta), h(\delta)+1\}$ (recall that $h(\delta)$ is the minimal element of $\mathcal{M}(\delta)$),

$$c(\delta) = \min\Big(\phi_\delta(h(\delta)-1) - \phi_\delta(h(\delta)), \phi_\delta(h(\delta)+2) - \phi_\delta(h(\delta)+1), \frac{1}{2}(\delta - \psi(\delta)), \frac{1}{2}\Big).$$

The minimum with $1/2$ and $(\delta - \psi(\delta))/2$ is taken for technical reasons since in later applications we will need $c(\delta)$ to be small enough, while (88) holds even without it. Note that the quantity $c(\delta)$ can be arbitrary close to 0. In fact, for any δ_0 such that $|\mathcal{M}(\delta_0)| = 2$, $c(\delta)$ is close to 0 if δ is close to δ_0.

Recall the notation $\bar{k}$ from (74). Now by the same chain of reasoning as in (80), setting $\xi := (2\varepsilon^{1/2} + 8\varepsilon\delta')^{1/4}$, we obtain that for some η_1, η_2 with $\lim_{\varepsilon\to 0}\eta_1 = \lim_{\varepsilon\to 0}\eta_2 = 0$,

$$\begin{aligned}
&\mathbb{P}(\bar{k} \notin \mathcal{M}(\delta),\ \lambda_1(Z^{(1)}) \geq \sqrt{2(1+\delta')\log n}) \\
&\leq C\log n \sum_{k\notin\mathcal{M}(\delta)} (d')^{\binom{k}{2}} n^{-\binom{k}{2}+k-\frac{k}{2(k-1)}(1-\xi)^2(1+\delta')+\frac{1+4\delta'}{2}\varepsilon^{1/2}+\varepsilon} \\
&+ Cn\cdot n^{-(1-\xi)^2(1+\delta')+\frac{1+4\delta'}{2}\varepsilon^{1/2}+\varepsilon} + Cn^{-2\delta'}
\end{aligned} \tag{89}$$

$$\leq Cn^{-\psi(\delta)-c(\delta)+\eta_1} + Cn^{-\delta+\eta_2}, \tag{90}$$

where the bound on the first term is obtained as follows. By (84), for each $k \notin \mathcal{M}(\delta)$, the exponent of n in (89) is bounded by

$$\begin{aligned} &-\binom{k}{2} + k - \frac{k}{2(k-1)}(1-\xi)^2(1+\delta') + \frac{1+4\delta'}{2}\varepsilon^{1/2} + \varepsilon \\ &\leq -\binom{k}{2} + k - \frac{k}{2(k-1)}(1+\delta) + r_\delta(\varepsilon) = -\phi_\delta(k) + r_\delta(\varepsilon) \overset{(88)}{\leq} -\psi(\delta) - c(\delta) + r_\delta(\varepsilon). \end{aligned}$$

Hence, by the argument (81)–(83), the term (89) can be bounded by $n^{-\psi(\delta)-c(\delta)+2r_\delta(\varepsilon)}$, and since $\lim_{\varepsilon\to 0} r_\delta(\varepsilon) = 0$, we obtain (90). Therefore, using the fact that $\psi(\delta) + c(\delta) < \delta$, for sufficiently small $\varepsilon > 0$,

$$\mathbb{P}(\bar{k} \notin \mathcal{M}(\delta),\ \lambda_1(Z^{(1)}) \geq \sqrt{2(1+\delta')\log n}) \leq Cn^{-\psi(\delta)-c(\delta)+\eta_1}. \tag{91}$$

Since the statement of the theorem is about the entire graph X and not just $X^{(1)}$, we will now show that superimposing $X^{(2)}$ on the latter does not alter the size of the maximal clique with high probability owing to the sparsity of $X^{(2)}$. Recall that we use k_X to denote the size of the maximal clique in X. Since $k_X \geq \bar{k}$ (recall that $\bar{k}$ is the maximal clique size in $X^{(1)}$), (91) implies

$$\mathbb{P}(k_X \leq h(\delta) - 1, \lambda_1(Z^{(1)}) \geq \sqrt{2(1+\delta')\log n}) \leq Cn^{-\psi(\delta)-c(\delta)+\eta_1}. \tag{92}$$

To treat the non-trivial direction, i.e., superimposing $X^{(2)}$ does not make k_X larger than $\bar{k}$, define the event $\mathcal{F}_1$, measurable with respect to $X^{(1)}$, by

$$\mathcal{F}_1 := \left\{ |E(H)| - \binom{\bar{k}}{2} \leq |V(H)| - \bar{k} : \text{ any subgraph } H \text{ in } X^{(1)} \text{ such that } |V(H)| \leq 2h(\delta) + 2 \right\}. \tag{93}$$

In words, under $\mathcal{F}_1$, the subgraph induced on any subset of vertices of size bigger than $\bar{k}$, has significantly smaller number of edges than the clique induced on the same subset.

Note that, in particular, if $\bar{k} \geq 4$, then on the event $\mathcal{F}_1$,

$$X^{(1)} \text{ has a unique maximal clique } K := K_{X^{(1)}} \text{ of size } \bar{k}. \tag{94}$$

This follows from the definition of $\mathcal{F}_1$ applied to the subgraph induced on $K \cup K'$ where K' is another set of $\bar{k}$ vertices.

We will show first show that $\mathcal{F}_1$ is likely, and on it, for X to have a larger clique, $X^{(2)}$ must fill in the 'substantially many' edges absent in $X^{(1)}$ which will then be shown to be unlikely.

Showing $\mathcal{F}_1$ is likely. Towards this, observe that

$$\mathbb{P}(\{\bar{k}=k\}\cap\mathcal{F}_1^c)\overset{(48)}{\leq} C\sum_{i=1}^{2h(\delta)+2}\left(\frac{i}{n}\right)^{\binom{k}{2}-k+1}\leq C(2h(\delta)+2)^{\binom{k}{2}-k+2}\frac{1}{n^{\binom{k}{2}-k+1}}. \quad (95)$$

Hence, recalling the event $\mathcal{F}_0$ in (77), using the above and the argument of (80) again, there is η_1' with $\lim_{\varepsilon\to 0}\eta_1'=0$ such that for $\delta>\delta_2$,

$$\begin{aligned}
&\mathbb{P}(\bar{k}\in\mathcal{M}(\delta),\mathcal{F}_1^c,\lambda_1(Z^{(1)})\geq\sqrt{2(1+\delta')\log n})\\
&\leq\sum_{k\in\mathcal{M}(\delta)}\mathbb{E}\left[\mathbb{P}(\max_{i\in I}\{\lambda_1(Z_i^{(1)})\}\geq\sqrt{2(1+\delta')\log n}\mid X^{(1)})\mathbb{1}_{\mathcal{F}_0}\mathbb{1}_{\mathcal{F}_1^c}\mathbb{1}_{\bar{k}=k}\right]\\
&+\mathbb{E}\left[\mathbb{P}(\max_{i\in J}\{\lambda_1(Z_i^{(1)})\}\geq\sqrt{2(1+\delta')\log n}\mid X^{(1)})\mathbb{1}_{\mathcal{D}_{4\delta'}\cap\mathcal{C}_{4\delta'}\cap\mathcal{E}_{4\delta'}}\right]+\mathbb{P}(\mathcal{F}_0^c)\\
&\leq C(\log n)n^{-1}\sum_{k\in\mathcal{M}(\delta)}(2h(\delta)+2)^{\binom{k}{2}-k+2}n^{-\binom{k}{2}+k-\frac{k}{2(k-1)}(1-\xi)^2(1+\delta')+\frac{1+4\delta'}{2}\varepsilon^{1/2}+\varepsilon}\\
&+Cn\cdot n^{-(1-\xi)^2(1+\delta')+\frac{1+4\delta'}{2}\varepsilon^{1/2}+\varepsilon}+Cn^{-2\delta'}\\
&\leq Cn^{-\psi(\delta)-1+\eta_1'}, \qquad (96)
\end{aligned}$$

where the extra n^{-1} factor in the first term comes from (95). Note that the condition $\delta>\delta_3$ implies that elements in $\mathcal{M}(\delta)$ are greater than or equal to 4, and hence the condition $k\geq 3$ in the summation throughout the argument in (80) is already implied by $k\in\mathcal{M}(\delta)$. Putting the above together, letting

$$\mathcal{F}_2:=\{\bar{k}\in\mathcal{M}(\delta)\}\cap\mathcal{F}_1, \quad (97)$$

by (91) and (96), for large δ,

$$\mathbb{P}(\mathcal{F}_2^c,\lambda_1(Z^{(1)})\geq\sqrt{2(1+\delta')\log n})\leq n^{-\psi(\delta)-c(\delta)+\eta_1} \quad (98)$$

(recall that $c(\delta)\in(0,1)$). By Lemma 5.2, this in particular implies

$$\begin{aligned}
&\mathbb{P}(\mathcal{F}_2^c,\lambda_1(Z)\geq\sqrt{2(1+\delta)\log n})\\
&\leq\mathbb{P}(\mathcal{F}_2^c,\lambda_1(Z^{(1)})\geq\sqrt{2(1+\delta')\log n})+\mathbb{P}(\lambda_1(Z)\geq\sqrt{2(1+\delta)\log n},\lambda_1(Z^{(1)})\\
&<\sqrt{2(1+\delta')\log n})\\
&\leq Cn^{-\psi(\delta)-c(\delta)+\eta_1}+\mathbb{P}(\lambda_1(Z^{(2)})\geq\sqrt{\varepsilon}(1+\delta)\sqrt{\log n})\leq Cn^{-\psi(\delta)-c(\delta)+\eta_1}. \qquad (99)
\end{aligned}$$

Combining this with (4), since $\lim_{\varepsilon\to 0}\eta_1=0$, there exists $\varepsilon_0=\varepsilon_0(\delta)>0$ such that for any $\varepsilon<\varepsilon_0$ (recall that ε implicitly appears in the definition of $X^{(1)}$),

$$\lim_{n\to\infty}\mathbb{P}(\mathcal{F}_2^c\mid\mathcal{U}_\delta)=0. \quad (100)$$

Since the minimum element in $\mathcal{M}(\delta)$ is greater than or equal to 4 under the condition $\delta > \delta_3$, the event $\mathcal{F}_2$ implies $\bar{k} \geq 4$ (see (97)). Hence, recalling $\mathcal{F}_2 \subset \mathcal{F}_1$ and the fact that $\mathcal{F}_1$ along with $\bar{k} \geq 4$ implies the uniqueness of maximal clique K in $X^{(1)}$ (see (94)),

$$\lim_{n\to\infty} \mathbb{P}(\text{there is a unique maximal clique } K \text{ in } X^{(1)} \mid \mathcal{U}_\delta) = 1. \tag{101}$$

For convenience, let us denote the above event by $Unique$. We now proceed to showing that the unique maximal clique K of $X^{(1)}$ continues to be so on superimposing $X^{(2)}$ to obtain X.

We first define some notations. For two subsets of vertices A and B, define the set of undirected edges

$$Edge(A, B) := \{e = (i, j) : i < j, i, j \in B\backslash A\} \cup \{e = (i, j) : i \in B\backslash A, j \in A \cap B\}.$$

Note that

$$|Edge(A, B)| = \binom{|B|}{2} - \binom{|A \cap B|}{2}. \tag{102}$$

Then, define the random subset of edges, measurable with respect to $X^{(1)}$, by

$$X^{(1)}(A, B) = Edge(A, B) \cap E(X^{(1)}).$$

We first verify that under the event $\mathcal{F}_2 = \{\bar{k} \in \mathcal{M}(\delta)\} \cap \mathcal{F}_1$, any clique K' of size $\ell \leq \bar{k}$ satisfies

$$|X^{(1)}(K, K')| \leq \ell - |K \cap K'| \tag{103}$$

where as mentioned above K is the unique maximal clique in $X^{(1)}$. Since

$$|E(K \cup K')| \geq \binom{\bar{k}}{2} + |X^{(1)}(K, K')|,$$

applying (93) to $H = K \cup K'$ (note that under the event $\bar{k} \in \mathcal{M}(\delta)$, we have $|K \cup K'| \leq 2\bar{k} \leq 2\,h(\delta) + 2$),

$$\binom{\bar{k}}{2} + |X^{(1)}(K, K')| - \binom{\bar{k}}{2} \leq |K \cup K'| - \bar{k} = \ell - |K \cap K'|,$$

which implies (103).

Note that conditioning on $X^{(1)}$, the entries of X are independent and satisfy

$$\mathbb{P}(X_{ij} = 1 | X^{(1)}_{ij} = 0) \leq \frac{2d}{n} \text{ for large } n,$$

$$\mathbb{P}(X_{ij} = 1 | X_{ij}^{(1)} = 1) = 1.$$

In fact, using the fact that $\mathbb{P}(X_{ij} = 0) = 1 - \frac{d}{n} \geq \frac{1}{2}$ for large n,

$$\mathbb{P}(X_{ij} = 1 | X_{ij}^{(1)} = 0) = \frac{\mathbb{P}(|Y_{ij}| < \sqrt{\varepsilon \log\log n}, X_{ij} = 1)}{\mathbb{P}(X_{ij}^{(1)} = 0)} \leq \frac{\mathbb{P}(X_{ij} = 1)}{\mathbb{P}(X_{ij} = 0)} \leq \frac{2d}{n},$$

and the second identity is obvious.

We will now define two events $\mathcal{B}_0$ and $\mathcal{B}_1$, which will be shown to be very likely on $\mathcal{U}_\delta$ and together would imply that K is the unique maximal clique in X and moreover, the largest clique not fully contained in K is a triangle.

We begin with $\mathcal{B}_0$ which is measurable with respect to the sigma algebra generated by $X^{(1)}$ and X.

$$\mathcal{B}_0 := (Unique \cap \{\text{there is a clique of size 4 in } X \text{ which is edge-disjoint from } K\})^c. \tag{104}$$

In other words, $\mathcal{B}_0^c$ demands a disjoint occurrence of the clique K and a clique of size 4 (in the graph X). Recalling that K is of size $\bar{k}$, by BK inequality and using Lemma 4.3

$$\mathbb{P}(\mathcal{B}_0^c \cap \{\bar{k} \in \mathcal{M}(\delta)\}) \leq C\left(\frac{1}{n}\right)^{\binom{h(\delta)}{2} - h(\delta)} \left(\frac{1}{n}\right)^{\binom{4}{2} - 4} = C\left(\frac{1}{n}\right)^{\binom{h(\delta)}{2} - h(\delta) + 2}, \tag{105}$$

where $C > 0$ is a constant depending only on δ. We write

$$\begin{aligned}
&\mathbb{P}\left(\mathcal{B}_0^c, \lambda_1(Z^{(1)}) \geq \sqrt{2(1+\delta')\log n}\right) \\
&\quad \leq \mathbb{E}\left[\mathbb{P}(\lambda_1(Z^{(1)}) \geq \sqrt{2(1+\delta')\log n} | X^{(1)}, X) \mathbb{1}_{\mathcal{F}_0} \mathbb{1}_{\bar{k} \in \mathcal{M}(\delta)} \mathbb{1}_{\mathcal{B}_0^c}\right] \\
&\qquad + \mathbb{P}\left(\left(\mathcal{F}_0 \cap \{\bar{k} \in \mathcal{M}(\delta)\}\right)^c, \lambda_1(Z^{(1)}) \geq \sqrt{2(1+\delta')\log n}\right).
\end{aligned} \tag{106}$$

Since $\lambda_1(Z^{(1)})$ and X are conditionally independent given $X^{(1)}$, by (75) and (76), there is $\eta_3 = \eta_3(\varepsilon)$ with $\lim_{\varepsilon \to 0} \eta_3 = 0$ such that for sufficiently small $\varepsilon > 0$,

$$\begin{aligned}
&\mathbb{P}(\lambda_1(Z^{(1)}) \geq \sqrt{2(1+\delta')\log n} | X^{(1)}, X) \mathbb{1}_{\mathcal{F}_0} \mathbb{1}_{\bar{k} \in \mathcal{M}(\delta)} \\
&\leq C(\log n) n^{-\frac{\bar{k}}{2(\bar{k}-1)}(1-\xi)^2(1+\delta') + \frac{1+4\delta'}{2}\varepsilon^{1/2} + \varepsilon} + Cn \cdot n^{-(1-\xi)^2(1+\delta') + \frac{1+4\delta'}{2}\varepsilon^{1/2} + \varepsilon} \\
&\leq Cn^{-\frac{\bar{k}}{2(\bar{k}-1)}(1+\delta) + \eta_3 + \varepsilon} \leq Cn^{-\frac{h(\delta)+1}{2h(\delta)}(1+\delta) + \eta_3 + \varepsilon},
\end{aligned}$$

where the second and last inequalities follow by observing $\frac{\bar{k}}{2(\bar{k}-1)}(1+\delta) \leq \phi_\delta(\bar{k}) < \phi_\delta(2) = \delta$ (since $\bar{k} \geq 3$ and $\delta > \delta_2$) and $\bar{k} \leq h(\delta)+1$ respectively. Hence, applying this and (105) to (106), using (78) and (98) to bound the last term in (106), for sufficiently

small $\varepsilon > 0$,

$$\begin{aligned}
&\mathbb{P}\left(\mathcal{B}_0^c, \lambda_1(Z^{(1)}) \geq \sqrt{2(1+\delta')\log n}\right) \\
&\leq Cn^{-\frac{h(\delta)+1}{2h(\delta)}(1+\delta)+\eta_3+\varepsilon}\mathbb{P}(\mathcal{B}_0^c \cap \{\bar{k} \in \mathcal{M}(\delta)\}) + Cn^{-2\delta'} + n^{-\psi(\delta)-c(\delta)+\eta_1} \\
&\overset{(105)}{\leq} Cn^{-\frac{h(\delta)+1}{2h(\delta)}(1+\delta)+\eta_3+\varepsilon}\left(\frac{1}{n}\right)^{\binom{h(\delta)}{2}-h(\delta)+2} + Cn^{-2\delta'} + n^{-\psi(\delta)-c(\delta)+\eta_1} \\
&\leq Cn^{-\psi(\delta)-1+\eta_3+\varepsilon} + Cn^{-2\delta'} + n^{-\psi(\delta)-c(\delta)+\eta_1} \leq Cn^{-\psi(\delta)-c(\delta)+\eta_1} \qquad (107)
\end{aligned}$$

(recall that $c(\delta) \in (0, 1)$), where the third inequality follows from the fact

$$\begin{aligned}
\frac{h(\delta)+1}{2h(\delta)}(1+\delta) + \binom{\tilde{h}(\delta)}{2} - h(\delta) &= \left(\frac{h(\delta)}{2(h(\delta)-1)}(1+\delta) + \frac{h(\delta)(h(\delta)-3)}{2}\right) \\
&\quad - \frac{1}{2h(\delta)(h(\delta)-1)} \\
&\geq \psi(\delta) - 1,
\end{aligned}$$

where the last inequality follows from the observation that the term in the parentheses is exactly $\psi(\delta)$. Let us define the another event, again measurable with respect to the sigma algebra generated by $X^{(1)}$ and X,

$$\mathcal{B}_1 := (Unique \cap \{\text{there is a clique } K' \text{ of size 4 in } X \text{ such that } 2 \leq |K \cap K'| \leq 3\})^c. \qquad (108)$$

Thus, in words, the event $\mathcal{B}_1^c$ demands the existence of a clique of size 4 which is not edge disjoint from K but also is not contained in the latter.

Note that by (102) and (103), under the event $\mathcal{F}_2$, the number of missing edges (of $X^{(1)}$) in $Edge(K, K')$ is

$$|Edge(K, K')\backslash X^{(1)}(K, K')| \geq \binom{|K'|}{2} - \binom{|K \cap K'|}{2} - (|K'| - |K \cap K'|).$$

Hence,

$$\begin{aligned}
&\mathbb{P}(\mathcal{B}_1^c \mid X^{(1)})\mathbb{1}_{\mathcal{F}_2} \\
&\leq \sum_{m=2}^{3} \sum_{|K\cap K'|=m} \mathbb{P}(X_{ij} = 1 \text{ for all edges } e \\
&= (i, j) \in Edge(K, K')\backslash X^{(1)}(K, K') \mid X^{(1)})\mathbb{1}_{\mathcal{F}_2} \\
&\leq \sum_{m=2}^{3} \bar{k}^m n^{4-m}\left(\frac{2d}{n}\right)^{\binom{4}{2}-\binom{m}{2}-(4-m)} \leq C\sum_{m=2}^{3}\left(\frac{1}{n}\right)^{(\binom{4}{2}-2\cdot 4)-(\binom{m}{2}-2m)} \leq C\frac{1}{n}.
\end{aligned} \qquad (109)$$

Hence, observing that, X and $\lambda_1(Z^{(1)})$ are conditionally independent given $X^{(1)}$, by (87) and (98),

$$\begin{aligned}
&\mathbb{P}\left(\mathcal{B}_1^c, \lambda_1(Z^{(1)}) \geq \sqrt{2(1+\delta')\log n}\right) \\
&\leq \mathbb{E}\left[\mathbb{P}(\mathcal{B}_1^c | X^{(1)}, \lambda_1(Z^{(1)}))\mathbb{1}_{\mathcal{F}_2}\mathbb{1}_{\lambda_1(Z^{(1)})\geq\sqrt{2(1+\delta')\log n}}\right] + \mathbb{P}(\mathcal{F}_2^c, \lambda_1(Z^{(1)}) \\
&\geq \sqrt{2(1+\delta')\log n}) \\
&\leq Cn^{-\psi(\delta)-c(\delta)+\eta_1}. \qquad (110)
\end{aligned}$$

Combining with (91) and (107),

$$\mathbb{P}\left((\mathcal{B}_0 \cap \mathcal{B}_1 \cap \{\bar{k} \in \mathcal{M}(\delta)\})^c, \lambda_1(Z^{(1)}) \geq \sqrt{2(1+\delta')\log n}\right) \leq Cn^{-\psi(\delta)-c(\delta)+\eta_1}. \qquad (111)$$

Proceeding as in (99)–(101), there exists $\varepsilon_1 > 0$ such that for any $\varepsilon < \varepsilon_1$,

$$\lim_{n\to\infty} \mathbb{P}(\mathcal{B}_0 \cap \mathcal{B}_1 \cap \{\bar{k} \in \mathcal{M}(\delta)\} \mid \mathcal{U}_\delta) = 1. \qquad (112)$$

Recall that conditionally on $\mathcal{U}_\delta$, the event $Unique$ happens with high probability (see (101)). Hence, by (112) and recalling the definition of $\mathcal{B}_0$ and $\mathcal{B}_1$, we have the following conclusion: With high probability conditionally on $\mathcal{U}_\delta$, there is a unique maximal clique K in $X^{(1)}$ and furthermore,

$$\text{there is no clique of size 4 in } X \text{ edge-disjoint from } K$$

and

$$\text{there is no clique } K' \text{ of size 4 in } X \text{ with } 2 \leq |K \cap K'| \leq 3.$$

This implies the statements in Theorem 1.4 and in particular

$$\lim_{n\to\infty} \mathbb{P}(\text{there is a unique maximal clique } K_X \text{ in } X \text{ and is equal to } K \mid \mathcal{U}_\delta) = 1. \qquad (113)$$

□

7 Optimal localization of leading eigenvector

We prove Theorem 1.7 in this section. Recall $v = (v_1, \cdots, v_n)$ is the unit eigenvector associated with the largest eigenvalue $\lambda_1 = \lambda_1(Z)$ and let K_X be the unique maximal clique (recall that Theorem 1.4 ensures uniqueness conditioned on $\mathcal{U}_\delta$ with high probability). Then,

$$\lambda_1 = \sum_{1\le i,j\le n} Z_{ij} v_i v_j = \sum_{1\le i,j\le n} Z^{(1)}_{ij} v_i v_j + \sum_{1\le i,j\le n} Z^{(2)}_{ij} v_i v_j. \tag{114}$$

The proof has two parts. In the first, we prove that the eigenvector allocates most of its mass on K_X, while in the second part we further show that the mass is uniformly distributed.

Mass concentration. Let us recall $r_\delta(\varepsilon)$ and $c(\delta)$ defined in (84) and (88) respectively. We choose a parameter ε sufficiently small so that

$$2r_\delta(\varepsilon) < c(\delta), \tag{115}$$

$$\varepsilon \le \frac{1}{\delta^4}, \tag{116}$$

$$\varepsilon < \min(\varepsilon_0, \varepsilon_1) \tag{117}$$

(ε_0 and ε_1 are positive constant depending on δ such that (101) and (113) are satisfied for $\varepsilon < \varepsilon_0$ and $\varepsilon < \varepsilon_1$ respectively). Recall that by (101) and (113), conditionally on $\mathcal{U}_\delta$, with probability tending to 1, the following is true: the maximal cliques $K_{X^{(1)}}$ and K_X are unique and equal which will be often denoted by K for brevity. Hence, throughout the proof, we assume the occurrence of this event.

Recall

$$\mathcal{A}_1 := \Big\{ \sum_{i\in K} v_i^2 \ge 1 - \kappa \Big\}, \tag{118}$$

where $\kappa > 0$ is the parameter in the statement of the theorem. Since

$$\mathbb{P}\left(\sum_{1\le i,j\le n} Z^{(2)}_{ij} v_i v_j \ge \sqrt{\varepsilon}(1+\delta)\sqrt{\log n} \right) \le \mathbb{P}\left(\lambda_1(X^{(2)}) \ge (1+\delta)\sqrt{\frac{\log n}{\log\log n}} \right) \le n^{-(2\delta+\delta^2)+o(1)},$$

by (65), for any event $\mathcal{A}$,

$$\mathbb{P}(\mathcal{A}, \lambda_1 \ge \sqrt{2(1+\delta)\log n}) \le \mathbb{P}(\mathcal{A}, \sum_{1\le i,j\le n} Z^{(1)}_{ij} v_i v_j \ge \sqrt{2(1+\delta')\log n}) + n^{-(2\delta+\delta^2)+o(1)}, \tag{119}$$

where $\delta' > 0$ as before is defined to be

$$\sqrt{2(1+\delta')} = \sqrt{2(1+\delta)} - \sqrt{\varepsilon}(1+\delta). \tag{120}$$

Note that since $\varepsilon \leq \frac{1}{\delta^4}$, using the bound for δ' in (67), we have

$$\delta' = \delta + o_\delta(1) \quad \text{as} \quad \delta \to \infty. \tag{121}$$

We will now bound the first term on the RHS of (119) with $\mathcal{A} = \mathcal{A}_1$ using Proposition 3.1 and the fact that on the high probability event $\mathcal{F}_1$ defined in (93), the largest clique outside K is at most a triangle (the reason will be explained shortly in (129) and the discussion following it) which would make it suboptimal in a large deviation theoretic sense for the eigenvector to allocate mass off of K. We now proceed to make this precise. The arguments will bear similarities with those appearing in the proof of Proposition 5.7.

Let $C_1, \cdots, C_m$ be connected components of $X^{(1)}$, and let without loss of generality C_1 contain the clique K of size $\bar{k}$. Let k_i be the maximum clique size in C_i.

We will now work with the high probability event $\mathcal{F}_0$ from (77). As in the proof of Proposition 5.7, define B_1 to be the collection of (directed) edges defined by

$$B_1 := \{e = (i, j) \in \overrightarrow{E(C_1)} : v_i^2, v_j^2 < \bar{\eta}^2\} \tag{122}$$

(recall that for any graph H, $\overrightarrow{E(H)}$ denotes the set of *directed* edges in H), where the parameter $\bar{\eta}$ is chosen to be

$$\bar{\eta} = \varepsilon^{1/4}. \tag{123}$$

Define the set of (directed) edges $B_2 := \overrightarrow{E(C_1)} \backslash B_1$. Since $|\{i : v_i^2 \geq \eta^2\}| \leq \frac{1}{\bar{\eta}^2}$, under the event $\mathcal{F}_0$, using the definition of $\mathcal{D}_{4\delta'}$,

$$\frac{1}{2}|B_2| \leq \frac{1 + 4\delta'}{\bar{\eta}^2} \frac{\log n}{\log \log n}, \tag{124}$$

by the same reasoning as preceding (54). We write

$$\sum_{(i,j)\in\overrightarrow{E(C_1)}} Z_{ij}^{(1)} v_i v_j = \sum_{(i,j)\in B_1} Z_{ij}^{(1)} v_i v_j + \sum_{(i,j)\in B_2} Z_{ij}^{(1)} v_i v_j =: S_1 + S_2. \tag{125}$$

By the same reasoning as in (58), and following the same notation as in the latter, for $2\bar{\eta}^2 \leq 1$, under the event $\mathcal{F}_0$,

$$\sum_{(i,j)\in B_1} v_i^2 v_j^2 \leq 2\bar{\eta}^2(1 - \bar{\eta}^2) + 2(4\delta' + 1)\bar{\eta}^4 =: \theta^4. \tag{126}$$

Note that since $\bar{\eta} = \varepsilon^{1/4} \leq \frac{1}{\delta}$ (see (116)), by (121), we have

$$\theta = O\Big(\frac{1}{\delta^{1/2}}\Big) \quad \text{as} \quad \delta \to \infty. \tag{127}$$

By Cauchy–Schwarz inequality and (126),

$$S_1 \le \Big(\sum_{(i,j)\in B_1} v_i^2 v_j^2 \Big)^{1/2} \Big(\sum_{(i,j)\in B_1} (Z_{ij}^{(1)})^2 \Big)^{1/2} \le \theta^2 \Big(\sum_{(i,j)\in \overrightarrow{E(C_1)}} (Z_{ij}^{(1)})^2 \Big)^{1/2}. \quad (128)$$

Next, we estimate S_2. Define $x := \sum_{i\in K} v_i^2$ and $y := \sum_{C_1\setminus K} v_i^2$. Recalling the definition of event $\mathcal{F}_1$ in (93), on the latter,

$$\text{the maximum size of clique in } K^c \text{ is at most 3.} \quad (129)$$

To see this, note that if K^c contains a clique K' of size 4, then $|E(K\cup K')| \ge \binom{\bar{k}}{2} + 6$ and $|V(K\cup K')| = \bar{k}+4$, which contradicts (93). Hence, under the event $\mathcal{F}_1$,

$$\begin{aligned} \sum_{(i,j)\in B_2} v_i^2 v_j^2 \le \sum_{(i,j)\in \overrightarrow{E(C_1)}} v_i^2 v_j^2 &\le \sum_{i\ne j, i,j\in K} v_i^2 v_j^2 + 2 \sum_{i\in K, j\in C_1\setminus K} v_i^2 v_j^2 \\ &\quad + \sum_{i\ne j,(i,j)\in \overrightarrow{E(C_1\setminus K)}} v_i^2 v_j^2 \\ &\le 2\Big(\frac{\bar{k}-1}{2\bar{k}} x^2 + xy + \frac{1}{3} y^2 \Big). \end{aligned} \quad (130)$$

The final bound follows from (21). Thus, under the event $\mathcal{F}_1$,

$$\begin{aligned} S_2 &\le \Big(\sum_{(i,j)\in B_2} v_i^2 v_j^2 \Big)^{1/2} \Big(\sum_{(i,j)\in B_2} (Z_{ij}^{(1)})^2 \Big)^{1/2} \\ &\le \Big(\frac{\bar{k}-1}{\bar{k}} x^2 + 2xy + \frac{2}{3} y^2 \Big)^{1/2} \Big(\sum_{(i,j)\in B_2} (Z_{ij}^{(1)})^2 \Big)^{1/2}. \end{aligned} \quad (131)$$

Note that using the fact

$$1 = \sum_{i=1}^n v_i^2 = x + y + \sum_{\ell=2}^m \Big(\sum_{i\in C_\ell} v_i^2 \Big),$$

we have

$$\begin{aligned} \sum_{1\le i,j\le n} Z_{ij}^{(1)} v_i v_j &= S_1 + S_2 + \sum_{\ell=2}^m \sum_{(i,j)\in \overrightarrow{E(C_\ell)}} Z_{ij}^{(1)} v_i v_j \\ &\le S_1 + S_2 + \sum_{\ell=2}^m \lambda(Z_\ell^{(1)}) \Big(\sum_{i\in C_\ell} v_i^2 \Big) \le S_1 + S_2 \\ &\quad + (1-x-y) \max_{\ell=2,\cdots,m} \lambda(Z_\ell^{(1)}). \end{aligned} \quad (132)$$

We now estimate the following conditional probability

$$\begin{aligned}
&\mathbb{P}\Big(x < 1-\kappa, \sum_{1\le i,j\le n} Z_{ij}^{(1)} v_i v_j \ge \sqrt{2(1+\delta')\log n} \mid X^{(1)}\Big)\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2} \\
&= \mathbb{P}\Big(x < 1-\kappa, y \ge \frac{\kappa}{2}, \sum_{1\le i,j\le n} Z_{ij}^{(1)} v_i v_j \ge \sqrt{2(1+\delta')\log n} \mid X^{(1)}\Big)\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2} \\
&\quad + \mathbb{P}\Big(x < 1-\kappa, y < \frac{\kappa}{2}, \sum_{1\le i,j\le n} Z_{ij}^{(1)} v_i v_j \ge \sqrt{2(1+\delta')\log n} \mid X^{(1)}\Big)\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2} \\
&=: R_1 + R_2. \qquad (133)
\end{aligned}$$

We next bound R_1 and R_2 in turn.

Bounding R_1. By (132),

$$\begin{aligned}
R_1 &\le \mathbb{P}(S_1 \ge \theta\sqrt{2(1+\delta')\log n} \mid X^{(1)})\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2} \\
&\quad + \mathbb{P}\Big(x < 1-\kappa, y \ge \frac{\kappa}{2}, S_2 \ge (x+y-\theta)\sqrt{2(1+\delta')\log n} \mid X^{(1)}\Big)\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2} \\
&\quad + \mathbb{P}\Big(\max_{\ell=2,\cdots,m} \lambda(Z_\ell^{(1)}) \ge \sqrt{2(1+\delta')\log n} \mid X^{(1)}\Big)\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2} \\
&=: R_{1,1} + R_{1,2} + R_{1,3}. \qquad (134)
\end{aligned}$$

By (128) and (39) in Lemma 5.1 with $\gamma = \varepsilon$, $L = \frac{1}{\theta^2}(1+\delta')\log n$ and $m \le \frac{2+4\delta'}{\varepsilon}\frac{\log n}{\log\log n} + 4\delta'$ (see (69) and (70)), for sufficiently large n,

$$R_{1,1} \le \mathbb{P}\Big(\sum_{i<j,(i,j)\in\overrightarrow{E(C_1)}} (Z_{ij}^{(1)})^2 \ge \frac{1}{\theta^2}(1+\delta')\log n \mid X^{(1)}\Big)\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2} \le n^{-\frac{1+\delta'}{2\theta^2}+(1+2\delta')+\varepsilon}. \qquad (135)$$

To bound $R_{1,2}$, we first need the following technical bound. There exists a constant $\lambda = \lambda(\kappa) \in (0, \frac{1}{100})$ such that for sufficiently large δ, under the event $\bar{k} \in \mathcal{M}(\delta)$, for $x < 1-\kappa$, $y \ge \frac{\kappa}{2}$,

$$\frac{\bar{k}-1}{\bar{k}}x^2 + 2xy + \frac{2}{3}y^2 < \Big(\frac{\bar{k}-1}{\bar{k}} - \lambda\Big)(x+y-\theta)^2. \qquad (136)$$

In fact, rearranging, this inequality holds if

$$\lambda x^2 + 2\Big(\frac{1}{\bar{k}} + \lambda\Big)xy + 2\Big(\frac{\bar{k}-1}{\bar{k}} - \lambda\Big)\theta(x+y) < \Big(\frac{\bar{k}-1}{\bar{k}} - \lambda - \frac{2}{3}\Big)y^2.$$

Recall that by (9), $\bar{k} \in \mathcal{M}(\delta)$ implies

$$\Big(\frac{1+\delta}{2}\Big)^{1/3} - 1 \le \bar{k} \le \Big(\frac{1+\delta}{2}\Big)^{1/3} + 3. \qquad (137)$$

Hence, there is $\lambda = \lambda(\kappa) > 0$ such that for $x < 1 - \kappa$, $y \geq \frac{\kappa}{2}$ and sufficiently large δ, under the event $\bar{k} \in \mathcal{M}(\delta)$,

$$\lambda x^2 < \frac{1}{10}y^2,\ 2\Big(\frac{1}{\bar{k}} + \lambda\Big)xy < \frac{1}{10}y^2,\ 2\Big(\frac{\bar{k}-1}{\bar{k}} - \lambda\Big)\theta(x+y) < 2\theta < \frac{1}{10}y^2$$

(see (127) for the bound of θ). If λ is small enough, say $\lambda \in (0, \frac{1}{100})$, then for sufficiently large δ, under the event $\bar{k} \in \mathcal{M}(\delta)$, $\frac{3}{10} < \frac{\bar{k}-1}{\bar{k}} - \lambda - \frac{2}{3}$, and thus we obtain (136).

Thus, by (131) and (136), using the fact $(\frac{\bar{k}-1}{\bar{k}} - \lambda)^{-1} \geq \frac{\bar{k}}{\bar{k}-1} + \lambda$,

$$R_{1,2} \leq \mathbb{P}\Big(\sum_{i<j,(i,j)\in B_2} (Z_{ij}^{(1)})^2 \geq \Big(\frac{\bar{k}}{\bar{k}-1} + \lambda\Big)(1+\delta')\log n \mid X^{(1)}\Big)\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2}. \quad (138)$$

Note that the event $\sum_{i<j,(i,j)\in B_2}(Z_{ij}^{(1)})^2 \geq t$ implies the existence of a random subset $J \in V(C_1)$ with $|J| \leq \Big\lfloor \frac{1}{\bar{\eta}^2} \Big\rfloor$ such that $\sum_{i<j,i \text{ or } j\in J, i\sim j} (Z_{ij}^{(1)})^2 \geq t$. Hence, by the union bound and (39) in Lemma 5.1 with

$$\gamma = \varepsilon,\ L = \Big(\frac{\bar{k}}{\bar{k}-1} + \lambda\Big)(1+\delta')\log n,\ m \leq \frac{1+4\delta'}{\bar{\eta}^2}\frac{\log n}{\log\log n}$$

(see (124)), recalling $\bar{\eta} = \varepsilon^{1/4}$, for large enough δ, for sufficiently large n,

$$R_{1,2} \leq |V(C_1)|^{\left\lfloor \frac{1}{\bar{\eta}^2}\right\rfloor} n^{-\frac{1}{2}(\frac{\bar{k}}{\bar{k}-1}+\lambda)(1+\delta')+\frac{\varepsilon}{2}\frac{1+4\delta'}{\bar{\eta}^2}+\varepsilon} \leq n^{-\frac{1}{2}\frac{\bar{k}}{\bar{k}-1}(1+\delta')-\frac{1}{2}\lambda(1+\delta')+\frac{1}{2}(1+4\delta')\varepsilon^{1/2}+2\varepsilon}. \quad (139)$$

Here, we used the fact $|V(C_1)| \leq (1+4\delta')\frac{\log n}{\log\log n}$ to bound the term $|V(C_1)|^{\left\lfloor \frac{1}{\bar{\eta}^2}\right\rfloor}$ by n^{ε}.

Recall from (129) that the size of maximal clique in C_ℓ, $\ell = 2, \cdots, m$, is at most 3 under the event $\mathcal{F}_2 = \{\bar{k} \in \mathcal{M}(\delta)\} \cap \mathcal{F}_1$. Hence, by Proposition 5.7 with

$$\alpha = 1+\delta',\ k \leq 3,\ \gamma = \varepsilon,\ \eta = \varepsilon^{1/4},\ c_1 = 1+4\delta',\ c_2 = \frac{2+4\delta'}{\varepsilon},\ c_3 = 4\delta',$$

setting $\xi := (2\eta^2 + 8\eta^4\delta')^{1/4}$, for sufficiently large δ,

$$\begin{aligned} R_{1,3} &\leq n(n^{-\frac{1+\delta'}{2\xi^2}+(1+2\delta')+\varepsilon} + n^{-\frac{3}{4}(1-\xi)^2(1+\delta')+\frac{1}{2}(1+4\delta')\varepsilon^{1/2}+\varepsilon}) \\ &\leq Cn^{-\frac{3}{4}(1-\xi)^2(1+\delta')+\frac{1}{2}(1+4\delta')\varepsilon^{1/2}+\varepsilon+1}. \end{aligned} \quad (140)$$

Here, we used the following comparison between exponents: as $\delta \to \infty$,

$$\frac{1+\delta'}{2\xi^2} - (1+2\delta') - \varepsilon = \Omega(\delta^2),$$
$$\frac{3}{4}(1-\xi)^2(1+\delta') - \frac{1}{2}(1+4\delta')\varepsilon^{1/2} - \varepsilon = \Big(\frac{3}{4} + o_\delta(1)\Big)\delta.$$

This follows from $\varepsilon \le \frac{1}{\delta^4}$ and the fact

$$\xi = O\Big(\frac{1}{\delta^{1/2}}\Big) \quad \text{as} \quad \delta \to \infty, \tag{141}$$

which is a consequence of $\xi = (2\eta^2 + 8\eta^4\delta')^{1/4} = (2\varepsilon^{1/2} + 8\varepsilon\delta')^{1/4}$, $\varepsilon \le \frac{1}{\delta^4}$ and the bound for δ' in (121).

Thus, applying the above bounds for $R_{1,1}$, $R_{1,2}$ and $R_{1,3}$ (see (135), (139) and (140) respectively) to (134), for sufficiently large δ,

$$R_1 \le Cn^{-\frac{1}{2}\frac{\bar{k}}{\bar{k}-1}(1+\delta') - \frac{1}{2}\lambda(1+\delta') + \frac{1}{2}(1+4\delta')\varepsilon^{1/2} + 2\varepsilon}. \tag{142}$$

This follows from the fact that for sufficiently large δ, under the event $\bar{k} \in \mathcal{M}(\delta)$, RHS of (139) is the slowest decaying term among itself, (135) and (140). In fact, using $\varepsilon \le \frac{1}{\delta^4}$ and the bound for θ, $\bar{k}$ and ξ in (127), (137) and (141) respectively, we have

$$\frac{1+\delta'}{2\theta^2} - (1+2\delta') - \varepsilon = \Omega(\delta^2), \tag{143}$$
$$\frac{1}{2}\frac{\bar{k}}{\bar{k}-1}(1+\delta') + \frac{1}{2}\lambda(1+\delta') - \frac{1}{2}(1+4\delta')\varepsilon^{1/2} - 2\varepsilon = \Big(\frac{1+\lambda}{2} + o_\delta(1)\Big)\delta, \tag{144}$$
$$\frac{3}{4}(1-\xi)^2(1+\delta') - \frac{1}{2}(1+4\delta')\varepsilon^{1/2} - \varepsilon - 1 = \Big(\frac{3}{4} + o_\delta(1)\Big)\delta. \tag{145}$$

Since $\lambda \in (0, \frac{1}{100})$, for large δ, (144) is smaller than the other two terms.

Bounding R_2. For $\upsilon > 0$ to be chosen later, we write

$$\begin{aligned}
R_2 &\le \mathbb{P}(S_1 \ge \theta\sqrt{2(1+\delta')\log n}|X^{(1)})\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2} \\
&\quad + \mathbb{P}\Big(S_2 \ge (1+\upsilon)\Big(x + \frac{\bar{k}}{\bar{k}-1}y\Big)\sqrt{2(1+\delta')\log n}|X^{(1)}\Big)\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2} \\
&\quad + \mathbb{P}\Big(x < 1-\kappa,\, y < \frac{\kappa}{2},\, (1-x-y)\max_{\ell\ge 2}\lambda_1(Z_\ell^{(1)})\Big) \\
&\ge \Big(1 - (1+\upsilon)\Big(x + \frac{\bar{k}}{\bar{k}-1}y\Big) - \theta\Big)\sqrt{2(1+\delta')\log n}|X^{(1)}\Big)\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2} \\
&=: R_{2,1} + R_{2,2} + R_{2,3}.
\end{aligned} \tag{146}$$

We take $\upsilon > 0$ such that for sufficiently large $\delta > 0$ and small $\kappa > 0$, under the event $\bar{k} \in \mathcal{M}(\delta)$, for any $x < 1 - \kappa$ and $y < \frac{\kappa}{2}$,

$$1 - (1+\upsilon)\Big(x + \frac{\bar{k}}{\bar{k}-1}y\Big) - \theta \ge 1 - (1+\upsilon)(x+y) - \frac{1+\upsilon}{2(\bar{k}-1)}\kappa - \theta$$
$$> \frac{9}{10}(1-x-y). \tag{147}$$

Here, the last inequality follows from the bound $x + y \le 1 - \frac{\kappa}{2}$ and the bound for θ in (127).

By (135), for sufficiently large n,

$$R_{2,1} \le n^{-\frac{1+\delta'}{2\theta^2}+(1+2\delta')+\varepsilon}. \tag{148}$$

Note that for sufficiently large δ, under the event $\bar{k} \in \mathcal{M}(\delta)$, by the bound for $\bar{k}$ in (137), we have $2(\frac{\bar{k}-1}{2\bar{k}}x^2 + xy + \frac{1}{3}y^2) < \frac{\bar{k}-1}{\bar{k}}(x + \frac{\bar{k}}{\bar{k}-1}y)^2$. Thus, by (131),

$$S_2 \le \Big(\frac{\bar{k}-1}{\bar{k}}\Big)^{1/2}\Big(x + \frac{\bar{k}}{\bar{k}-1}y\Big)\Big(\sum_{(i,j)\in B_2}(Z_{ij}^{(1)})^2\Big)^{1/2}.$$

Hence, by the same arguments as in (138) and (139) (apply (39) in Lemma 5.1 with $\gamma = \varepsilon$, $L = \frac{\bar{k}}{\bar{k}-1}(1+\upsilon)^2(1+\delta')\log n$ and $m \le \frac{1+4\delta'}{\tilde{\eta}^2}\frac{\log n}{\log\log n}$), for large enough δ, for sufficiently large n,

$$\begin{aligned} R_{2,2} &\le \mathbb{P}\Big(\sum_{i<j,(i,j)\in B_2}(Z_{ij}^{(1)})^2 \ge \frac{\bar{k}}{\bar{k}-1}(1+\upsilon)^2(1+\delta')\log n \mid X^{(1)}\Big)\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2} \\ &\le |V(C_1)|^{\left\lfloor\frac{1}{\tilde{\eta}^2}\right\rfloor} n^{-\frac{1}{2}\frac{\bar{k}}{\bar{k}-1}(1+\upsilon)^2(1+\delta')+\frac{1}{2}(1+4\delta')\varepsilon^{1/2}+\varepsilon} \\ &\le n^{-\frac{1}{2}\frac{\bar{k}}{\bar{k}-1}(1+\delta')-\upsilon(1+\delta')+\frac{1}{2}(1+4\delta')\varepsilon^{1/2}+2\varepsilon}. \end{aligned} \tag{149}$$

Next, by (147),

$$R_{2,3} \le \mathbb{P}\Big(\max_{\ell=2,\cdots,m}\lambda_1(Z_\ell^{(1)}) \ge \frac{9}{10}\sqrt{2(1+\delta')\log n} \mid X^{(1)}\Big)\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2}.$$

Since the size of maximal clique in C_ℓ, $\ell = 2, \cdots, m$, is at most 3 under the event $\mathcal{F}_2$, by the same argument as in (140) (apply Proposition 5.7 with $\alpha = (\frac{9}{10})^2(1+\delta')$, $k \le 3$, $\eta = \varepsilon^{1/4}$ and $\gamma = \varepsilon$), for sufficiently large δ,

$$\begin{aligned} R_{2,3} &\le n(n^{-(\frac{9}{10})^2\frac{1+\delta'}{2\xi^2}+(1+2\delta')+\varepsilon} + n^{-\frac{3}{4}(\frac{9}{10})^2(1+\delta')+\frac{1}{2}(1+4\delta')\varepsilon^{1/2}+\varepsilon}) \\ &\le Cn^{-\frac{3}{4}(\frac{9}{10})^2(1+\delta')+\frac{1}{2}(1+4\delta')\varepsilon^{1/2}+\varepsilon+1}. \end{aligned} \tag{150}$$

Thus, applying (148), (149) and (150) to (146), for large enough δ, for sufficiently large n,

$$R_2 \le Cn^{-\frac{1}{2}\frac{\bar{k}}{\bar{k}-1}(1+\delta')-\upsilon(1+\delta')+\frac{1}{2}(1+4\delta')\varepsilon^{1/2}+2\varepsilon}. \tag{151}$$

This follows from the fact that for sufficiently large δ, under the event $\bar{k} \in \mathcal{M}(\delta)$, RHS of (149) is the slowest decaying term among itself, (148) and (150). This can be verified by the similar argument as in (142), combined with the fact that

$$\frac{1}{2}\frac{\bar{k}}{\bar{k}-1}(1+\delta')+\upsilon(1+\delta')-\frac{1}{2}(1+4\delta')\varepsilon^{1/2}-2\varepsilon = \Big(\frac{1}{2}+\upsilon+o_\delta(1)\Big)\delta$$

and $\frac{1}{2} < \frac{3}{4}(\frac{9}{10})^2$.

Therefore, using the bounds in (142) and (151) in (133) we get that

$$\begin{aligned}\mathbb{P}\Big(x < 1-\kappa, \sum_{1\le i,j\le n} Z_{ij}^{(1)} v_i v_j \ge \sqrt{2(1+\delta')\log n} \mid X^{(1)}\Big)\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2} \\ \le Cn^{-\frac{1}{2}\frac{\bar{k}}{\bar{k}-1}(1+\delta')-\min(\frac{1}{2}\lambda,\upsilon)(1+\delta')+\frac{1}{2}(1+4\delta')\varepsilon^{1/2}+2\varepsilon}.\end{aligned} \tag{152}$$

Finishing the proof. Recall that under the event $\mathcal{F}_0$, the number of non-tree components is less than $\log n$. Hence, by (75) and (76),

$$\begin{aligned}&\mathbb{P}\Big(\sum_{1\le i,j\le n} Z_{ij}^{(1)} v_i v_j \ge \sqrt{2(1+\delta')\log n} \mid X^{(1)}\Big)\mathbb{1}_{\mathcal{F}_0} \\ &\le \mathbb{P}\Big(\max_{\ell=1,\cdots,m} \lambda_1(Z_\ell^{(1)}) \ge \sqrt{2(1+\delta')\log n} \mid X^{(1)}\Big)\mathbb{1}_{\mathcal{F}_0} \\ &\le C(\log n)n^{-\frac{\bar{k}}{2(\bar{k}-1)}(1-\xi)^2(1+\delta')+\frac{1}{2}(1+4\delta')\varepsilon^{1/2}+\varepsilon} + Cn\cdot n^{-(1-\xi)^2(1+\delta')+\frac{1}{2}(1+4\delta')\varepsilon^{1/2}+\varepsilon}\end{aligned} \tag{153}$$

(recall that $\xi = (2\varepsilon^{1/2}+8\varepsilon\delta')^{1/4}$). Using the bound for ξ in (141) and $\varepsilon \le \frac{1}{\delta^4}$, in the case $\bar{k} \ge 3$, for large δ, for sufficiently large n, (153) is bounded by

$$Cn^{-\frac{\bar{k}}{2(\bar{k}-1)}(1-\xi)^2(1+\delta')+\frac{1}{2}(1+4\delta')\varepsilon^{1/2}+2\varepsilon}, \tag{154}$$

and for $\bar{k} = 2$, (153) is bounded by

$$Cn^{-(1-\xi)^2(1+\delta')+\frac{1}{2}(1+4\delta')\varepsilon^{1/2}+\varepsilon+1}. \tag{155}$$

Recalling $\mathcal{F}_2 = \{\bar{k} \in \mathcal{M}(\delta)\} \cap \mathcal{F}_1$, we write

$$\mathbb{P}\Big(x < 1-\kappa, \sum_{1\le i,j\le n} Z_{ij}^{(1)} v_i v_j \ge \sqrt{2(1+\delta')\log n}\Big)$$

$$\le \sum_{k\in\mathcal{M}(\delta)} \mathbb{E}\Big[\mathbb{P}\Big(x<1-\kappa, \sum_{1\le i,j\le n} Z_{ij}^{(1)} v_i v_j \ge \sqrt{2(1+\delta')\log n} \mid X^{(1)}\Big)\mathbb{1}_{\bar{k}=k}\mathbb{1}_{\mathcal{F}_0}\mathbb{1}_{\mathcal{F}_2}\Big]$$
$$+ \sum_{k\in\mathcal{M}(\delta)} \mathbb{E}\Big[\mathbb{P}\Big(\sum_{1\le i,j\le n} Z_{ij}^{(1)} v_i v_j \ge \sqrt{2(1+\delta')\log n} \mid X^{(1)}\Big)\mathbb{1}_{\bar{k}=k}\mathbb{1}_{\mathcal{F}_0}\mathbb{1}_{\mathcal{F}_1^c}\Big]$$
$$+ \sum_{k\notin\mathcal{M}(\delta)} \mathbb{E}\Big[\mathbb{P}\Big(\sum_{1\le i,j\le n} Z_{ij}^{(1)} v_i v_j \ge \sqrt{2(1+\delta')\log n} \mid X^{(1)}\Big)\mathbb{1}_{\bar{k}=k}\mathbb{1}_{\mathcal{F}_0}\Big] + \mathbb{P}(\mathcal{F}_0^c). \quad (156)$$

Recalling $\varepsilon \le \frac{1}{\delta^4}$, by (79) and (152), the first term in (156) is bounded by

$$C\sum_{k\in\mathcal{M}(\delta)} n^{-\frac{1}{2}\frac{k}{k-1}(1+\delta')-\min(\frac{1}{2}\lambda,\upsilon)(1+\delta')+\frac{1}{2}(1+4\delta')\varepsilon^{1/2}+2\varepsilon}\frac{(d')_{(2)}^{(k)}}{n^{\binom{k}{2}-k}}. \quad (157)$$

We use the argument in (81)–(83) to bound this quantity. The exponent in n above is less than

$$\Big[-\binom{k}{2}+k-\frac{k}{2(k-1)}(1-\xi)^2(1+\delta')+\frac{1+4\delta'}{2}\varepsilon^{1/2}+2\varepsilon\Big]-\min\Big(\frac{1}{2}\lambda,\upsilon\Big)(1+\delta').$$

Comparing this with the exponent in (81), we notice the additional term $\min(\frac{1}{2}\lambda,\upsilon)(1+\delta')$. Hence, recalling η_1 in (83) can be chosen as $\eta_1 = 2r_\delta(\varepsilon)$ (see (85)), (157) can be bounded by

$$Cn^{-\psi(\delta)-\min(\frac{1}{2}\lambda,\upsilon)(1+\delta')+2r_\delta(\varepsilon)}. \quad (158)$$

Similarly, using (95) and (154), the second term in (156) is bounded by

$$C\sum_{k\in\mathcal{M}(\delta)} n^{-\frac{k}{2(k-1)}(1-\xi)^2(1+\delta')+\frac{1}{2}(1+4\delta')\varepsilon^{1/2}+2\varepsilon}\Big(\frac{2h(\delta)+2}{n}\Big)^{\binom{k}{2}-k+1} \le Cn^{-\psi(\delta)-1+2r_\delta(\varepsilon)}. \quad (159)$$

This follows from the fact that there is an additional n^{-1} term arising from $(\frac{2h(\delta)+2}{n})^{\binom{k}{2}-k+1}$. In addition, by (79), (154) and (155), the third term in (156) is bounded by

$$C\sum_{k\ge 3,k\notin\mathcal{M}(\delta)} n^{-\frac{\bar{k}}{2(\bar{k}-1)}(1-\xi)^2(1+\delta')+\frac{1}{2}(1+4\delta')\varepsilon^{1/2}+2\varepsilon}\frac{(d')_{(2)}^{(k)}}{n^{\binom{k}{2}-k}} + Cn^{-(1-\xi)^2(1+\delta')+\frac{1}{2}(1+4\delta')\varepsilon^{1/2}+\varepsilon+1}$$
$$\le Cn^{-\psi(\delta)-c(\delta)+2r_\delta(\varepsilon)}. \quad (160)$$

Here, the additional term $c(\delta)$ comes from the fact that in the first term, the summation is taken only over $k\in\mathcal{M}(\delta)^c$ and $\phi_\delta(k)\ge\psi(\delta)+c(\delta)$ for $k\in\mathcal{M}(\delta)^c$ (see (90) for details). The second term can be absorbed in the constant C since $\psi(\delta)=(\frac{1}{2}+o_\delta(1))\delta$ (see (6)).

Finally, by (78), the last term in (156) is bounded by $n^{-2\delta'}$. Hence, applying the above bounds to (156), for sufficiently large $\delta > 0$,

$$\mathbb{P}\Big(x < 1-\kappa, \sum_{1\le i,j\le n} Z_{ij}^{(1)} v_i v_j \ge \sqrt{2(1+\delta')\log n}\Big) \le C n^{-\psi(\delta)-c(\delta)+2r_\delta(\varepsilon)}. \quad (161)$$

Above, we used the fact that for large enough δ, $c(\delta) < 1 < \min(\frac{1}{2}\lambda, \upsilon)(1+\delta')$, which follows from the bound for δ' in (121). Since $2r_\delta(\varepsilon) < c(\delta)$ (see (115)), applying (119) and then Theorem 1.1, for sufficiently large δ,

$$\lim_{n\to\infty} \mathbb{P}(x < 1-\kappa \mid \mathcal{U}_\delta) = 0. \quad (162)$$

Therefore, for sufficiently large δ,

$$\lim_{n\to\infty} \mathbb{P}(\mathcal{A}_1 \mid \mathcal{U}_\delta) = 1. \quad (163)$$

□

Uniformity of eigenvector. We will aim to show $\sum_{i<j,i,j\in K}(v_i^2 - v_j^2)^2$ is small from which the form of uniformity appearing in the theorem statement follows immediately.

We first recall the parameter θ defined in (126). By (132), setting $\rho := 16\kappa$,

$$\begin{aligned}
&\mathbb{P}\Big(x \ge 1-\kappa, \sum_{i<j,i,j\in K}(v_i^2-v_j^2)^2 > \rho, \sum_{1\le i,j\le n} Z_{ij}^{(1)} v_i v_j \ge \sqrt{2(1+\delta')\log n} \mid X^{(1)}\Big)\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2} \\
&\le \mathbb{P}(S_1 \ge \theta\sqrt{2(1+\delta')\log n} \mid X^{(1)})\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2} \\
&\quad + \mathbb{P}\Big(x \ge 1-\kappa, \sum_{i<j,i,j\in K}(v_i^2-v_j^2)^2 > \rho, S_2 \ge (x+y-\theta)\sqrt{2(1+\delta')\log n} \mid X^{(1)}\Big)\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2} \\
&\quad + \mathbb{P}\Big(\max_{\ell=2,\cdots,m} \lambda(Z_\ell^{(1)}) \ge \sqrt{2(1+\delta')\log n} \mid X^{(1)}\Big)\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2}. \quad (164)
\end{aligned}$$

Since the first and third terms were already estimated during the analysis in the first part, we now estimate the second term. Using the identity

$$\frac{\bar{k}-1}{2\bar{k}}\Big(\sum_{i\in K} v_i^2\Big)^2 - \sum_{i<j,i,j\in K} v_i^2 v_j^2 = \frac{1}{2\bar{k}} \sum_{i<j,i,j\in K} (v_i^2 - v_j^2)^2,$$

under the event $\mathcal{F}_1$, we have an improvement of (130):

$$\begin{aligned}
\sum_{(i,j)\in B_2} v_i^2 v_j^2 &\le \sum_{i\ne j,i,j\in K} v_i^2 v_j^2 + 2\sum_{i\in K, j\in C_1\setminus K} v_i^2 v_j^2 + \sum_{i\ne j,(i,j)\in\overrightarrow{E(C_1\setminus K)}} v_i^2 v_j^2 \\
&\le 2\Big(\frac{\bar{k}-1}{2\bar{k}}x^2 + xy + \frac{1}{3}y^2\Big) - \frac{1}{\bar{k}} \sum_{i<j,i,j\in K} (v_i^2 - v_j^2)^2
\end{aligned}$$

where we used the above identity to bound the first term on the RHS. Thus, under the event $\sum_{i<j,i,j\in K}(v_i^2 - v_j^2)^2 > \rho$, using $x \le 1$, we obtain the analog of (131):

$$S_2 \le \Big(\Big(\frac{\bar{k}-1}{\bar{k}} - \frac{\rho}{\bar{k}}\Big)x^2 + 2xy + \frac{2}{3}y^2\Big)^{1/2}\Big(\sum_{(i,j)\in B_2}(Z_{ij}^{(1)})^2\Big)^{1/2}. \tag{165}$$

To bound the above, we need the following technical inequality. For sufficiently large δ, under the event $\bar{k} \in \mathcal{M}(\delta)$, for $x \ge 1-\kappa$,

$$\Big(\frac{\bar{k}-1}{\bar{k}} - \frac{\rho}{\bar{k}}\Big)x^2 + 2xy + \frac{2}{3}y^2 < \Big(\frac{\bar{k}-1}{\bar{k}} - \frac{\rho}{2\bar{k}}\Big)(x+y-\theta)^2. \tag{166}$$

In fact, by rearranging, (166) holds for sufficiently large δ if

$$2\Big(\frac{1}{\bar{k}} + \frac{\rho}{2\bar{k}}\Big)xy + 2\Big(\frac{\bar{k}-1}{\bar{k}} - \frac{\rho}{2\bar{k}}\Big)\theta(x+y) \le \frac{\rho}{2\bar{k}}x^2, \tag{167}$$

since using (137) we know the coefficient of y^2 on the RHS is at least that on the LHS.

For $x \ge 1-\kappa$, we have $y \le \kappa$ and thus $2(\frac{1}{\bar{k}} + \frac{\rho}{2\bar{k}})xy \le \frac{\rho}{4\bar{k}}x^2$ holds for small enough $\kappa > 0$ (recall $\rho = 16\kappa$). Also, by the bounds for θ and $\bar{k}$ in (127) and (137) respectively, under the event $\bar{k} \in \mathcal{M}(\delta)$, we have $2(\frac{\bar{k}-1}{\bar{k}} - \frac{\rho}{2\bar{k}})\theta(x+y) \le 2\theta \le \frac{\rho}{4\bar{k}}x^2$. The previous two inequalities imply (167) and thus (166).

Hence, by (165) and (166), and further using $(\frac{\bar{k}-1}{\bar{k}} - \frac{\rho}{2\bar{k}})^{-1} \ge \frac{\bar{k}}{\bar{k}-1} + \frac{\rho}{2\bar{k}}$, the second term in (164) is bounded by

$$\mathbb{P}\Big(\sum_{i<j,(i,j)\in B_2}(Z_{ij}^{(1)})^2 \ge \Big(\frac{\bar{k}}{\bar{k}-1} + \frac{\rho}{2\bar{k}}\Big)(1+\delta')\log n \mid X^{(1)}\Big)\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2}. \tag{168}$$

As before, by the union bound and Lemma 5.1 with $\gamma = \varepsilon$, $L = (\frac{\bar{k}}{\bar{k}-1} + \frac{\rho}{2\bar{k}})(1+\delta')\log n$ and $m \le \frac{1+4\delta'}{\bar{\eta}^2}\frac{\log n}{\log\log n}$, the above, and thus the second term in (164), is bounded by

$$\begin{aligned}|V(C_1)|^{\lfloor \frac{1}{\bar{\eta}^2}\rfloor} n^{-\frac{1}{2}(\frac{\bar{k}}{\bar{k}-1}+\frac{\rho}{2\bar{k}})(1+\delta')+\frac{1}{2}(1+4\delta')\varepsilon^{1/2}+\varepsilon}\\ \le n^{-\frac{1}{2}\frac{\bar{k}}{\bar{k}-1}(1+\delta')-\frac{\rho}{4\bar{k}}(1+\delta')+\frac{1}{2}(1+4\delta')\varepsilon^{1/2}+2\varepsilon}.\end{aligned} \tag{169}$$

Since the first and last terms in (164) are bounded by (135) and (140) respectively, one can deduce that

$$\begin{aligned}\mathbb{P}\Big(x \ge 1-\kappa, \sum_{i<j,i,j\in K}(v_i^2 - v_j^2)^2 > \rho, \sum_{1\le i,j\le n} Z_{ij}^{(1)} v_i v_j \ge \sqrt{2(1+\delta')\log n}\,|X^{(1)}\Big)\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2}\\ \le Cn^{-\frac{1}{2}\frac{\bar{k}}{\bar{k}-1}(1+\delta')-\frac{\rho}{4\bar{k}}(1+\delta')+\frac{1}{2}(1+4\delta')\varepsilon^{1/2}+2\varepsilon}.\end{aligned} \tag{170}$$

Similarly as (156), we write

$$\begin{aligned}
&\mathbb{P}\Big(x \ge 1-\kappa, \sum_{i<j,i,j\in K}(v_i^2-v_j^2)^2 > \rho, \sum_{1\le i,j\le n} Z_{ij}^{(1)} v_i v_j \ge \sqrt{2(1+\delta')\log n}\Big) \\
&\le \sum_{k\in\mathcal{M}(\delta)} \mathbb{E}\Big[\mathbb{P}\Big(x \ge 1-\kappa, \sum_{i<j,i,j\in K}(v_i^2-v_j^2)^2 > \rho, \sum_{1\le i,j\le n} Z_{ij}^{(1)} v_i v_j \\
&\ge \sqrt{2(1+\delta')\log n} \mid X^{(1)}\Big)\mathbb{1}_{\bar{k}=k}\mathbb{1}_{\mathcal{F}_0}\mathbb{1}_{\mathcal{F}_2}\Big] \\
&+ \sum_{k\in\mathcal{M}(\delta)} \mathbb{E}\Big[\mathbb{P}\Big(\sum_{1\le i,j\le n} Z_{ij}^{(1)} v_i v_j \ge \sqrt{2(1+\delta')\log n} \mid X^{(1)}\Big)\mathbb{1}_{\bar{k}=k}\mathbb{1}_{\mathcal{F}_0}\mathbb{1}_{\mathcal{F}_1^c}\Big] \\
&+ \sum_{k\notin\mathcal{M}(\delta)} \mathbb{E}\Big[\mathbb{P}\Big(\sum_{1\le i,j\le n} Z_{ij}^{(1)} v_i v_j \ge \sqrt{2(1+\delta')\log n} \mid X^{(1)}\Big)\mathbb{1}_{\bar{k}=k}\mathbb{1}_{\mathcal{F}_0}\Big] + \mathbb{P}(\mathcal{F}_0^c).
\end{aligned} \tag{171}$$

Using (137) and (170), there exist a constant $c > 0$ such that the first term in (171) is bounded by

$$C\sum_{k\in\mathcal{M}(\delta)} n^{-\frac{1}{2}\frac{\bar{k}}{\bar{k}-1}(1+\delta')-\frac{\rho}{4\bar{k}}(1+\delta')+\frac{1}{2}(1+4\delta')\varepsilon^{1/2}+2\varepsilon}\frac{(d')_2^{(k)}}{n_2^{(k)}-k} \le Cn^{-\psi(\delta)-c\rho\delta^{2/3}+2r_\delta(\varepsilon)}. \tag{172}$$

Other three terms in (171) can be bounded using (159), (160) and (78) respectively. Hence, combining these together, using the fact that $c(\delta) < 1 < c\rho\delta^{2/3}$ for large δ, we have

$$\begin{aligned}
&\mathbb{P}\Big(x \ge 1-\kappa, \sum_{i<j,i,j\in K}(v_i^2-v_j^2)^2 > \rho, \sum_{1\le i,j\le n} Z_{ij}^{(1)} v_i v_j \\
&\ge \sqrt{2(1+\delta')\log n}\Big) \le Cn^{-\psi(\delta)-c(\delta)+2r_\delta(\varepsilon)}.
\end{aligned}$$

Since $2r_\delta(\varepsilon) < c(\delta)$, applying (119) and then Theorem 1.1,

$$\lim_{n\to\infty}\mathbb{P}\Big(x \ge 1-\kappa, \sum_{i<j,i,j\in K}(v_i^2-v_j^2)^2 > \rho \mid \mathcal{U}_\delta\Big) = 0,$$

and thus by (162),

$$\lim_{n\to\infty}\mathbb{P}\Big(x \ge 1-\kappa, \sum_{i<j,i,j\in K}(v_i^2-v_j^2)^2 \le \rho \mid \mathcal{U}_\delta\Big) = 1.$$

It is now straightforward to obtain the uniformity statement in the theorem from the smallness of $\sum_{i<j,i,j\in K}(v_i^2 - v_j^2)^2$. To see this, note that setting $S := \sum_{i\in K} v_i^2$,

$$\sum_{i\in K}\Big(v_i^2 - \frac{1}{|K|}S\Big)^2 = \sum_{i\in K} v_i^4 - 2\frac{S}{|K|}\sum_{i\in K} v_i^2 + \frac{1}{|K|}S^2 = \sum_{i\in K} v_i^4 - \frac{1}{|K|}S^2$$
$$= \frac{1}{|K|}\Big((K-1)\sum_{i\in K} v_i^4 - 2\sum_{i<j,i,j\in K} v_i^2 v_j^2\Big) = \frac{1}{|K|}\sum_{i<j,i,j\in K}(v_i^2 - v_j^2)^2. \tag{173}$$

Hence, recalling $\rho = 16\kappa$, for sufficiently large δ,

$$\lim_{n\to\infty}\mathbb{P}\Big(x \ge 1-\kappa, \sum_{i\in K}\Big(v_i^2 - \frac{1}{|K|}\sum_{i\in K} v_i^2\Big)^2 \le \frac{16\kappa}{|K|} \mid \mathcal{U}_\delta\Big) = 1. \tag{174}$$

Using the inequality $(a+b)^2 \le 2(a^2+b^2)$, under the event

$$\Big\{\sum_{i\in K} v_i^2 \ge 1-\kappa\Big\} \cap \Big\{\sum_{i\in K}\Big(v_i^2 - \frac{1}{|K|}\sum_{i\in K} v_i^2\Big)^2 \le \frac{16\kappa}{|K|}\Big\},$$

we have

$$\sum_{i\in K}\Big(v_i^2 - \frac{1}{|K|}\Big)^2 \le 2\sum_{i\in K}\Big(v_i^2 - \frac{1}{|K|}\sum_{i\in K} v_i^2\Big)^2 + 2|K|\Big(\frac{1}{|K|}\sum_{i\in K} v_i^2 - \frac{1}{|K|}\Big)^2$$
$$\le \frac{32\kappa}{|K|} + \frac{2\kappa^2}{|K|} \le \frac{40\kappa}{|K|} =: \frac{\kappa_0}{|K|}. \tag{175}$$

Recalling that $K_X = K$, the proof is complete.

8 Uniform largeness of Gaussian weights

We prove Theorem 1.6 in this section. The proof essentially proceeds by comparing the ℓ_1 and ℓ_2 norms of the Gaussian variables on the edges of the clique K_X by obtaining sharp estimates on each of them. The final statement then can be deduced from a quantitative version of the Cauchy–Schwarz inequality. However, as the statement of the theorem indicates, we will end up working with a set T slightly smaller than K_X. Implementing the strategy involves a few steps and in particular relies on Theorem 1.7 which is the reason we proved the latter first.

Sum of squares of the Gaussian weights. We use the same notations as in Sect. 7. Also, as in the beginning of the proof of Theorem 1.7, we assume that the maximal cliques $K := K_{X^{(1)}}$ and K_X are unique and equal.

Setting $\rho := 16\kappa$, similarly as (166), for sufficiently large δ, under the event $\bar{k} \in \mathcal{M}(\delta)$, for $x \geq 1-\kappa$,

$$\frac{\bar{k}-1}{\bar{k}}x^2 + 2xy + \frac{2}{3}y^2 \leq \Big(\frac{\bar{k}-1}{\bar{k}} + \frac{\rho}{\bar{k}}\Big)(x+y-\theta)^2. \tag{176}$$

Using the above and (131),

$$S_2 \leq \Big(\frac{\bar{k}-1}{\bar{k}} + \frac{\rho}{\bar{k}}\Big)^{1/2}(x+y-\theta)\Big(\sum_{(i,j)\in B_2}(Z_{ij}^{(1)})^2\Big)^{1/2}, \tag{177}$$

where S_1 and S_2 were defined in (125). We now define an event guaranteeing a sharp behavior of the ℓ_2 norm of the Gaussian variables on the edges in B_2 where the latter was defined below (122),

$$\mathcal{A}_3 := \Big\{2\Big(\frac{\bar{k}}{\bar{k}-1} - \frac{\rho}{\bar{k}}\Big)(1+\delta')\log n \leq \sum_{(i,j)\in B_2}(Z_{ij}^{(1)})^2 \leq 2\Big(\frac{\bar{k}}{\bar{k}-1} + \frac{\rho}{\bar{k}}\Big)(1+\delta')\log n\Big\}. \tag{178}$$

Thus we have

$$\begin{aligned}
&\mathbb{P}\Big(\mathcal{A}_3^c, x \geq 1-\kappa, \sum_{1\leq i,j\leq n} Z_{ij}^{(1)} v_i v_j \geq \sqrt{2(1+\delta')\log n} \mid X^{(1)}\Big)\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2} \\
&\leq \mathbb{P}(S_1 \geq \theta\sqrt{2(1+\delta')\log n} \mid X^{(1)})\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2} \\
&\quad + \mathbb{P}\Big(\mathcal{A}_3^c, x \geq 1-\kappa, S_2 \geq (x+y-\theta)\sqrt{2(1+\delta')\log n} \mid X^{(1)}\Big)\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2} \\
&\quad + \mathbb{P}\Big(\max_{\ell=2,\cdots,m}\lambda(Z_\ell^{(1)}) \geq \sqrt{2(1+\delta')\log n} \mid X^{(1)}\Big)\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2}.
\end{aligned} \tag{179}$$

Since the first and last terms above can be bounded using (135) and (140) respectively, we only bound the second term.

-Bounding the second term: Using $(\frac{\bar{k}-1}{\bar{k}} + \frac{\rho}{\bar{k}})^{-1} \geq \frac{\bar{k}}{\bar{k}-1} - \frac{\rho}{\bar{k}}$,

$$\begin{aligned}
&\mathbb{P}\Big(\mathcal{A}_3^c, x \geq 1-\kappa, S_2 \geq (x+y-\theta)\sqrt{2(1+\delta')\log n} \mid X^{(1)}\Big)\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2} \\
&\overset{(177)}{\leq} \mathbb{P}\Big(\mathcal{A}_3^c, \sum_{(i,j)\in B_2}(Z_{ij}^{(1)})^2 \geq 2\Big(\frac{\bar{k}}{\bar{k}-1} - \frac{\rho}{\bar{k}}\Big)(1+\delta')\log n \mid X^{(1)}\Big)\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2} \\
&\leq \mathbb{P}\Big(\sum_{i<j,(i,j)\in B_2}(Z_{ij}^{(1)})^2 \geq \Big(\frac{\bar{k}}{\bar{k}-1} + \frac{\rho}{\bar{k}}\Big)(1+\delta')\log n \mid X^{(1)}\Big)\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2},
\end{aligned} \tag{180}$$

where the last inequality follows from the definition of $\mathcal{A}_3$. As before, by union bound and (39) in Lemma 5.1 with $\gamma = \varepsilon$, $L = (\frac{\bar{k}}{\bar{k}-1} + \frac{\rho}{\bar{k}})(1+\delta')\log n$ and $m \leq \frac{1+4\delta'}{\eta^2}\frac{\log n}{\log\log n}$

(using the bound on $|B_2|$ in (124)), for sufficiently large n, the above, and thus the second term in (179), is bounded by

$$|V(C_1)|^{\left\lfloor \frac{1}{\eta^2}\right\rfloor} n^{-\frac{1}{2}(\frac{\bar{k}}{\bar{k}-1}+\frac{\rho}{\bar{k}})(1+\delta')+\frac{1}{2}(1+4\delta')\varepsilon^{1/2}+\varepsilon} \le n^{-\frac{1}{2}\frac{\bar{k}}{\bar{k}-1}(1+\delta')-\frac{\rho}{2\bar{k}}(1+\delta')+\frac{1}{2}(1+4\delta')\varepsilon^{1/2}+2\varepsilon}. \tag{181}$$

-Combining altogether: As mentioned above, the first and last terms in (179) can be bounded using (135) and (140) respectively. Hence, combining these together,

$$\mathbb{P}\Big(\mathcal{A}_3^c, x \ge 1-\kappa, \sum_{1\le i,j\le n} Z_{ij}^{(1)} v_i v_j \ge \sqrt{2(1+\delta')\log n} \mid X^{(1)}\Big)\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2} \le Cn^{-\frac{1}{2}\frac{\bar{k}}{\bar{k}-1}(1+\delta')-\frac{\rho}{2\bar{k}}(1+\delta')+\frac{1}{2}(1+4\delta')\varepsilon^{1/2}+2\varepsilon}. \tag{182}$$

This follows from the fact that for sufficiently large δ, under the event $\mathcal{F}_0\cap\mathcal{F}_2$, (181) is the slowest decaying term among itself, (135) and (140). This follows from (143) and (145) and observing that $\varepsilon \le \frac{1}{\delta^4}$ and the bound for $\bar{k}$ in (137) together, under the event $\bar{k}\in\mathcal{M}(\delta)$, implies

$$\frac{1}{2}\frac{\bar{k}}{\bar{k}-1}(1+\delta')+\frac{\rho}{2\bar{k}}(1+\delta')-\frac{1}{2}(1+4\delta')\varepsilon^{1/2}-2\varepsilon = \Big(\frac{1}{2}+o_\delta(1)\Big)\delta.$$

Similarly as in (156), we write

$$\begin{aligned}
&\mathbb{P}\Big(\mathcal{A}_3^c, x\ge 1-\kappa, \sum_{1\le i,j\le n} Z_{ij}^{(1)} v_i v_j \ge \sqrt{2(1+\delta')\log n}\Big)\\
&\le \sum_{k\in\mathcal{M}(\delta)} \mathbb{E}\Big[\mathbb{P}\Big(\mathcal{A}_3^c, x\ge 1-\kappa, \sum_{1\le i,j\le n} Z_{ij}^{(1)} v_i v_j \ge \sqrt{2(1+\delta')\log n}\mid X^{(1)}\Big)\mathbb{1}_{\bar{k}=k}\mathbb{1}_{\mathcal{F}_0}\mathbb{1}_{\mathcal{F}_2}\Big]\\
&\quad + \sum_{k\in\mathcal{M}(\delta)} \mathbb{E}\Big[\mathbb{P}\Big(\sum_{1\le i,j\le n} Z_{ij}^{(1)} v_i v_j \ge \sqrt{2(1+\delta')\log n}\mid X^{(1)}\Big)\mathbb{1}_{\bar{k}=k}\mathbb{1}_{\mathcal{F}_0}\mathbb{1}_{\mathcal{F}_1^c}\Big]\\
&\quad + \sum_{k\notin\mathcal{M}(\delta)} \mathbb{E}\Big[\mathbb{P}\Big(\sum_{1\le i,j\le n} Z_{ij}^{(1)} v_i v_j \ge \sqrt{2(1+\delta')\log n}\mid X^{(1)}\Big)\mathbb{1}_{\bar{k}=k}\mathbb{1}_{\mathcal{F}_0}\Big] + \mathbb{P}(\mathcal{F}_0^c).
\end{aligned} \tag{183}$$

First, as in (157), one can bound the first term above using (182). In fact, using the bound for δ' and $\bar{k}$ in (121) and (137) respectively, under $\bar{k}\in\mathcal{M}(\delta)$, $\frac{\rho}{2\bar{k}}(1+\delta') \ge c\delta^{2/3}$ for some $c>0$. Thus, for large δ, the first term in (183) can be bounded by $Cn^{-\psi(\delta)-c'\delta^{2/3}}$ for some $c'<c$. Combining this with the bounds for other three terms, previously obtained in (159), (160) and (78) respectively, using the fact that $c(\delta)<1<c'\delta^{2/3}$ for large δ, (183) is bounded by $Cn^{-\psi(\delta)-c(\delta)+2r_\delta(\varepsilon)}$. Hence, using (119), for large δ,

$$\mathbb{P}(\mathcal{A}_3^c, x\ge 1-\kappa, \lambda_1 \ge \sqrt{2(1+\delta)\log n}) \le Cn^{-\psi(\delta)-c(\delta)+2r_\delta(\varepsilon)}. \tag{184}$$

Since $2r_\delta(\varepsilon) < c(\delta)$, combined with Theorem 1.1 and (162), for sufficiently large δ,

$$\lim_{n\to\infty} \mathbb{P}\Big(\mathcal{A}_3 \mid \mathcal{U}_\delta\Big) = 1. \tag{185}$$

Sum of absolute values of Gaussian weights. We now estimate the sum of absolute values of $Z_{ij}^{(1)}$. Defining

$$\mathcal{A}' := \Big\{ \sum_{i\in K} \Big(v_i^2 - \frac{1}{\bar{k}}\Big)^2 \le \frac{\kappa_0}{\bar{k}} \Big\} \tag{186}$$

(recall $\kappa_0 = 40\kappa$, see (175)), since $\bar{k} = |K|$, by (174) and (175),

$$\lim_{n\to\infty} \mathbb{P}\Big(\mathcal{A}' \mid \mathcal{U}_\delta\Big) = 1. \tag{187}$$

Recalling that $K = K_X$ with probability going to one conditionally on $\mathcal{U}_\delta$, the events $\mathcal{A}_2$ (from the statement of the theorem) and $\mathcal{A}'$ are essentially the same.

We now define the set of vertices T appearing in the statement of the theorem,

$$T := \Big\{ i \in K : \Big| v_i^2 - \frac{1}{\bar{k}} \Big| < \frac{\kappa_0^{1/4}}{\bar{k}} \Big\}.$$

Then, by (137), for sufficiently large δ, under the event $|\bar{k} - h(\delta)| \le 1$,

$$i \ne j, i, j \in T \quad \text{implies} \quad (i, j) \in B_2. \tag{188}$$

This is because for $i \in T$ and large δ, $v_i^2 \ge (1 - \kappa_0^{1/4})\frac{1}{\bar{k}} \overset{(137)}{\ge} \frac{c}{\delta^{1/3}} > \frac{1}{\delta^2} \ge \bar{\eta}^2$ where the final inequality is by our choice of $\bar{\eta}$ in (123). We now write

$$\begin{aligned} S_2 = \sum_{(i,j)\in B_2} Z_{ij}^{(1)} v_i v_j &= \sum_{i \text{ or } j\in T^c, (i,j)\in B_2} Z_{ij}^{(1)} v_i v_j + \sum_{i,j\in T, (i,j)\in B_2} Z_{ij}^{(1)} v_i v_j \\ &=: S_{21} + S_{22} \end{aligned} \tag{189}$$

(see (125) for the definition of S_2). By Cauchy–Schwarz inequality, under the event $\mathcal{A}'$,

$$\sum_{i\in K} \Big| v_i^2 - \frac{1}{\bar{k}} \Big| \le \kappa_0^{1/2}. \tag{190}$$

Thus, under the event $\mathcal{A}'$,

$$|T^c \cap K| \le \kappa_0^{1/4}\bar{k}, \qquad |T| \ge (1 - \kappa_0^{1/4})\bar{k}. \tag{191}$$

Note that (190) implies $\sum_{v_i^2 \ge \frac{1}{\bar{k}}(1+\kappa_0^{1/4})}(v_i^2 - \frac{1}{\bar{k}}) \le \kappa_0^{1/2}$, and thus under the event $\mathcal{A}'$,

$$\sum_{i\in T^c} v_i^2 = \sum_{v_i^2 \ge \frac{1}{\bar{k}}(1+\kappa_0^{1/4})} v_i^2 + \sum_{v_i^2 \le \frac{1}{\bar{k}}(1-\kappa_0^{1/4})} v_i^2 \le \Big(\kappa_0^{1/2} + \frac{1}{\bar{k}}\kappa_0^{1/4}\bar{k}\Big) + \frac{1}{\bar{k}}\kappa_0^{1/4}\bar{k} = \kappa_0^{1/2} + 2\kappa_0^{1/4}.$$

Hence, under the event $\mathcal{A}'$,

$$\sum_{i \text{ or } j\in T^c,(i,j)\in B_2} v_i^2 v_j^2 \le 2\Big(\sum_{i\in T^c} v_i^2\Big)\Big(\sum_{j\in C_1} v_j^2\Big) \le 2\kappa_0^{1/2} + 4\kappa_0^{1/4} =: \kappa'^2, \tag{192}$$

and thus

$$\begin{aligned} S_{21} &\le \Big(\sum_{i \text{ or } j\in T^c,(i,j)\in B_2} (Z_{ij}^{(1)})^2\Big)^{1/2}\Big(\sum_{i \text{ or } j\in T^c,(i,j)\in B_2} v_i^2 v_j^2\Big)^{1/2} \\ &\le \kappa'\Big(\sum_{i \text{ or } j\in T^c,(i,j)\in B_2} (Z_{ij}^{(1)})^2\Big)^{1/2}. \end{aligned} \tag{193}$$

In addition, using the fact that $v_i^2 < \frac{1}{\bar{k}}(1+\kappa_0^{1/4})$ for $i \in T$,

$$|S_{22}| \le \frac{1}{\bar{k}}(1+\kappa_0^{1/4}) \sum_{i,j\in T,(i,j)\in B_2} |Z_{ij}^{(1)}| \le \frac{1}{\bar{k}}(1+\kappa_0^{1/4}) \sum_{i\ne j, i,j\in T} |Z_{ij}^{(1)}|. \tag{194}$$

Now, we define the following event analogous to $\mathcal{A}_3$, but for the ℓ_1 norm,

$$\mathcal{A}_4 := \Big\{\bar{k}(1-3\kappa^{1/4})\sqrt{2(1+\delta')\log n} \le \sum_{i\ne j, i,j\in T} |Z_{ij}^{(1)}| \le \bar{k}(1+3\kappa^{1/4})\sqrt{2(1+\delta')\log n}\Big\}. \tag{195}$$

Now using the decomposition in (132) and further using (189),

$$\sum_{1\le i,j\le n} Z_{ij}^{(1)} v_i v_j \le S_1 + S_{21} + S_{22} + (1-x-y) \max_{\ell=2,\cdots,m} \lambda(Z_\ell^{(1)}),$$

we write

$$\begin{aligned} &\mathbb{P}\Big(\mathcal{A}_4^c, \mathcal{A}', x \ge 1-\kappa, \sum_{1\le i,j\le n} Z_{ij}^{(1)} v_i v_j \ge \sqrt{2(1+\delta')\log n} \mid X^{(1)}\Big)\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2} \\ &\quad\le \mathbb{P}(S_1 \ge \theta\sqrt{2(1+\delta')\log n}|X^{(1)})\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2} + \mathbb{P}(\mathcal{A}', S_{21} \\ &\quad\ge \sqrt{\kappa'}\sqrt{2(1+\delta')\log n} \mid X^{(1)})\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2} \\ &\qquad + \mathbb{P}\Big(\mathcal{A}_4^c, x \ge 1-\kappa, S_{22} \ge (x+y-\theta-\sqrt{\kappa'})\sqrt{2(1+\delta')\log n} \mid X^{(1)}\Big)\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2} \\ &\qquad + \mathbb{P}\Big(\max_{\ell=2,\cdots,m} \lambda(Z_\ell^{(1)}) \ge \sqrt{2(1+\delta')\log n} \mid X^{(1)}\Big)\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2} \end{aligned} \tag{196}$$

(recall that κ' is defined in (192)). Since we already have estimates for the first and last terms above, we only focus on the second and third terms.

-Bounding the second term: By (193),

$$\begin{aligned}&\mathbb{P}(\mathcal{A}', S_{21} \geq \sqrt{\kappa'}\sqrt{2(1+\delta')\log n} \mid X^{(1)})\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2}\\ &\leq \mathbb{P}\Big(\mathcal{A}', \sum_{i<j,\ i \text{ or } j\in T^c, (i,j)\in B_2} (Z_{ij}^{(1)})^2 \geq \frac{1}{\kappa'}(1+\delta')\log n \mid X^{(1)}\Big)\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2}. \quad (197)\end{aligned}$$

Note that by (39) in Lemma 5.1 with

$$\gamma = \varepsilon,\ L = \frac{1}{\kappa'}(1+\delta')\log n,\ m \leq 4\kappa^{1/4}\bar{k}(1+4\delta')\frac{\log n}{\log\log n}$$

(see (68)), for sufficiently large n, the quantity (197), and thus the second term in (196), is bounded by

$$|V(C_1)|^{\lfloor 4\kappa^{1/4}\bar{k}\rfloor} n^{-\frac{1}{2\kappa'}(1+\delta')+\frac{1}{2}4\kappa^{1/4}\bar{k}(1+4\delta')\varepsilon+\varepsilon} \leq n^{-\frac{1}{2\kappa'}(1+\delta')+2\kappa^{1/4}\bar{k}(1+4\delta')\varepsilon+2\varepsilon}. \quad (198)$$

The above inequality follows from the bound for $|V(C_1)|$ in (69) and observing that

$$\Big(\frac{2+4\delta'}{\varepsilon}\frac{\log n}{\log\log n}\Big)^{\lfloor 4\kappa^{1/4}\bar{k}\rfloor} \overset{(137)}{\leq} \Big(\frac{2+4\delta'}{\varepsilon}\frac{\log n}{\log\log n}\Big)^{c\delta^{1/3}} \leq n^{\varepsilon}$$

for large n ($c > 0$ is a constant depending on κ). The first factor in (198), as several times before, appears due to a union bound over all possible choices of $T^c \cap K$.
-Bounding the third term: Note that for sufficiently small $\kappa > 0$, for large enough δ and $x \geq 1-\kappa$,

$$1 - 3\kappa^{1/4} \leq \frac{x+y-\theta-\sqrt{\kappa'}}{1+\kappa_0^{1/4}} \quad (199)$$

(recall $\kappa_0 = 40\kappa$). In fact, (199) holds if $(1-3\kappa^{1/4})(1+\kappa_0^{1/4}) \leq 1-\kappa-\theta-\sqrt{\kappa'}$ for sufficiently large δ and small $\kappa > 0$, which follows from the bound for θ in (127).

Hence, using (194) and (199), recalling the definition of $\mathcal{A}_4$ in (195), the third term in (196) can be controlled by

$$\begin{aligned}&\mathbb{P}\Big(\mathcal{A}_4^c, x \geq 1-\kappa, S_{22} \geq (x+y-\theta-\sqrt{\kappa'})\sqrt{2(1+\delta')\log n} \mid X^{(1)}\Big)\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2}\\ &\leq \mathbb{P}\Big(\mathcal{A}_4^c, \sum_{i\neq j, i, j\in T} |Z_{ij}^{(1)}| \geq \bar{k}(1-3\kappa^{1/4})\sqrt{2(1+\delta')\log n} \mid X^{(1)}\Big)\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2}\\ &\leq \mathbb{P}\Big(\sum_{i\neq j, i, j\in T} |Z_{ij}^{(1)}| > \bar{k}(1+3\kappa^{1/4})\sqrt{2(1+\delta')\log n} \mid X^{(1)}\Big)\mathbb{1}_{\mathcal{F}_0\cap\mathcal{F}_2}\end{aligned}$$

$$\le \mathbb{P}\Big(\sum_{i<j, i, j \in T} (Z_{ij}^{(1)})^2 > \frac{\bar{k}^2}{\bar{k}(\bar{k}-1)} (1+3\kappa^{1/4})^2 (1+\delta') \log n \mid X^{(1)} \Big) \mathbb{1}_{\mathcal{F}_0 \cap \mathcal{F}_2}. \tag{200}$$

The second inequality follows from the definition of $\mathcal{A}_4$, (similar to (180)), while in the third inequality above, we used the Cauchy–Schwarz inequality and the fact $|T| \le \bar{k}$. Hence, by the union bound and (39) in Lemma 5.1 with $\gamma = \varepsilon$, $L = \frac{\bar{k}}{\bar{k}-1}(1+3\kappa^{1/4})^2(1+\delta') \log n$ and $m \le \bar{k}^2$, for sufficiently large n, the quantity (200), and thus the third term in (196), is bounded by

$$|V(C_1)|^{\bar{k}} n^{-\frac{1}{2}\frac{\bar{k}}{\bar{k}-1}(1+3\kappa^{1/4})^2(1+\delta')+\varepsilon} \le n^{-\frac{1}{2}\frac{\bar{k}}{\bar{k}-1}(1+3\kappa^{1/4})^2(1+\delta')+2\varepsilon}. \tag{201}$$

Here, we used the bound for $|V(C_1)|$ in (69) and the upper bound for $\bar{k}$ in (137) under the event $\bar{k} \in \mathcal{M}(\delta)$.

-*Combining altogether:* As mentioned already, the first and the last terms in (196) can be bounded by (135) and (140) respectively. Thus, combining these with (198) and (201), for sufficiently small $\kappa > 0$ and large δ, for large enough n,

$$\begin{aligned} \mathbb{P}\Big(\mathcal{A}_4^c, \mathcal{A}', x \ge 1-\kappa, \sum_{1 \le i, j \le n} Z_{ij}^{(1)} v_i v_j \ge \sqrt{2(1+\delta') \log n} \mid X^{(1)} \Big) \mathbb{1}_{\mathcal{F}_0 \cap \mathcal{F}_2} \\ \le C n^{-\frac{1}{2}\frac{\bar{k}}{\bar{k}-1}(1+3\kappa^{1/4})^2(1+\delta')+2\varepsilon} \le C n^{-\frac{1}{2}\frac{\bar{k}}{\bar{k}-1}(1+\delta')-3\kappa^{1/4}\delta'+2\varepsilon}. \end{aligned} \tag{202}$$

This follows from the fact that for sufficiently small $\kappa > 0$ and large δ, under the event $\mathcal{F}_0 \cap \mathcal{F}_2$, (201) is the slowest decaying term among itself, (135), (140) and (198). In fact, using $\varepsilon \le \frac{1}{\delta^4}$ and the bound for $\bar{k}$ in (137), under the event $\bar{k} \in \mathcal{M}(\delta)$,

$$\begin{aligned} \frac{1}{2\kappa'}(1+\delta') - 2\kappa^{1/4}\bar{k}(1+4\delta')\varepsilon - 2\varepsilon &= \Big(\frac{1}{2\kappa'} + o_\delta(1) \Big) \delta, \\ \frac{1}{2}\frac{\bar{k}}{\bar{k}-1}(1+3\kappa^{1/4})^2(1+\delta') - 2\varepsilon &= \Big(\frac{1}{2}(1+3\kappa^{1/4})^2 + o_\delta(1) \Big) \delta. \end{aligned}$$

Hence, recalling the definition of κ' in (192), for small $\kappa > 0$ and large δ, the quantity (201) slowly decays than (198). Also, by comparing the above asymptotic with (143) and (145), one can deduce that the quantity (201) slowly decays than (135) and (140) for small κ and large δ.

Thus, by proceeding as in (156), for sufficiently large δ,

$$\mathbb{P}\Big(\mathcal{A}_4^c, \mathcal{A}', x \ge 1-\kappa, \sum_{1 \le i, j \le n} Z_{ij}^{(1)} v_i v_j \ge \sqrt{2(1+\delta') \log n} \Big) \le n^{-\psi(\delta)-c(\delta)+2r_\delta(\varepsilon)}.$$

In fact, one can bound this quantity by the sum of four quantities via the argument of (156). Using (202) and the bound for $\bar{k}$ in (137), for large δ, the corresponding first

term in (156) can be bounded by $Cn^{-\psi(\delta)-c\delta}$ for some $c > 0$, and other three terms can be bounded by (159), (160) and (78) respectively. Combining these together, using the fact that $c(\delta) < 1 < c\delta$ for large δ, we obtain the above inequality. Applying (119) and then Theorem 1.1,

$$\lim_{n\to\infty} \mathbb{P}\Big(\mathcal{A}_4^c, \mathcal{A}', x \geq 1-\kappa \mid \mathcal{U}_\delta\Big) = 0.$$

Hence, by (162) and (187),

$$\lim_{n\to\infty} \mathbb{P}\Big(\mathcal{A}_4 \mid \mathcal{U}_\delta\Big) = 1. \tag{203}$$

Finishing the proof. Finally, using (185) and (203), we finish the proof. Define the event

$$\mathcal{A}_5 := \{\bar{k} \in \mathcal{M}(\delta)\}.$$

By (100), recalling $\mathcal{A}_5 \subset \mathcal{F}_2$,

$$\lim_{n\to\infty} \mathbb{P}\Big(\mathcal{A}_5 \mid \mathcal{U}_\delta\Big) = 1. \tag{204}$$

Now, define the event

$$\mathcal{A}_6 := \mathcal{A}' \cap \mathcal{A}_3 \cap \mathcal{A}_4 \cap \mathcal{A}_5.$$

Since $\mathcal{A}'$, $\mathcal{A}_3$, $\mathcal{A}_4$ and $\mathcal{A}_5$ are typical events conditioned on $\mathcal{U}_\delta$ (see (187), (185), (203) and (204) respectively),

$$\lim_{n\to\infty} \mathbb{P}\Big(\mathcal{A}_6 \mid \mathcal{U}_\delta\Big) = 1. \tag{205}$$

We next verify that the event $\mathcal{A}_6$ implies the desired uniformity of Gaussians claimed in the statement of the theorem. As indicated earlier, the proof involves technical manipulations involving the Cauchy–Schwarz inequality to relate the ℓ_1 and ℓ_2 norms.

For the ease of reading, let us recall the events

$$\mathcal{A}_3 = \Big\{2\Big(\frac{\bar{k}}{\bar{k}-1} - \frac{\rho}{\bar{k}}\Big)(1+\delta')\log n \leq \sum_{(i,j)\in B_2} (Z_{ij}^{(1)})^2 \leq 2\Big(\frac{\bar{k}}{\bar{k}-1} + \frac{\rho}{\bar{k}}\Big)(1+\delta')\log n\Big\},$$

$$\mathcal{A}_4 = \Big\{\bar{k}(1-3\kappa^{1/4})\sqrt{2(1+\delta')\log n} \leq \sum_{i\neq j, i, j\in T} |Z_{ij}^{(1)}| \leq \bar{k}(1+3\kappa^{1/4})\sqrt{2(1+\delta')\log n}\Big\}.$$

Note that using the fact $\rho = 16\kappa$ and (188), for sufficiently large δ, the event $\mathcal{A}_3 \cap \mathcal{A}_4 \cap \mathcal{A}_5$ implies that

$$\frac{1}{2} \sum_{i\neq j, i'\neq j', i, j, i', j'\in T} (|Z_{ij}^{(1)}| - |Z_{i'j'}^{(1)}|)^2$$

$$
\begin{aligned}
&= |T|(|T|-1)\Big(\sum_{i\neq j, i,j\in T}(Z_{ij}^{(1)})^2\Big) - \Big(\sum_{i\neq j,i,j\in T}|Z_{ij}^{(1)}|\Big)^2 \\
&\le 2|T|(|T|-1)\Big(\frac{\bar{k}}{\bar{k}-1}+\frac{\rho}{\bar{k}}\Big)(1+\delta')\log n - 2\bar{k}^2(1-3\kappa^{1/4})^2(1+\delta')\log n \\
&\le (32(\bar{k}-1)\kappa + 12\kappa^{1/4}\bar{k}^2)(1+\delta')\log n \le C\kappa^{1/4}\bar{k}^2(1+\delta')\log n,
\end{aligned}
$$

where we used $|T| \le \bar{k}$ and $\kappa \le \kappa^{1/4}$ in the second and the last inequality respectively. From this, using the argument in (173), setting $S' = \sum_{i,j\in T}|Z_{ij}^{(1)}|$, one can deduce that

$$
\sum_{i\neq j,i,j\in T}\Big(|Z_{ij}^{(1)}| - \frac{1}{|T|(|T|-1)}S'\Big)^2 \le C\kappa^{1/4}(1+\delta')\log n. \tag{206}
$$

We check that under the event $\mathcal{A}'\cap\mathcal{A}_4\cap\mathcal{A}_5$, there exists $\iota(\kappa)$ with $\lim_{\kappa\to 0}\iota = 0$ such that

$$
\Big|\frac{1}{|T|(|T|-1)}S' - \frac{1}{h(\delta)}\sqrt{2(1+\delta')\log n}\Big| \le \Big(\iota(\kappa)+\frac{C}{h(\delta)}\Big)\frac{1}{h(\delta)}\sqrt{2(1+\delta')\log n}. \tag{207}
$$

In fact, first note that by (191), $|T| \ge (1-\kappa_0^{1/4})\bar{k}$ under the event $\mathcal{A}'$. Also, $\bar{k} \ge h(\delta)$ under $\mathcal{A}_5$ and we have the upper bound for S' under $\mathcal{A}_4$. Hence, combining these ingredients together, under the event $\mathcal{A}'\cap\mathcal{A}_4\cap\mathcal{A}_5$,

$$
\begin{aligned}
\frac{1}{|T|(|T|-1)}S' &\le \frac{1+3\kappa^{1/4}}{1-\kappa_0^{1/4}}\frac{1}{\bar{k}(1-\kappa_0^{1/4})-1}\sqrt{2(1+\delta')\log n} \\
&\le (1+10\kappa^{1/4})\frac{1}{h(\delta)(1-\kappa_0^{1/4})-1}\sqrt{2(1+\delta')\log n} \\
&\le (1+10\kappa^{1/4})\Big(1+2\kappa_0^{1/4}+\frac{2}{h(\delta)}\Big)\frac{1}{h(\delta)}\sqrt{2(1+\delta')\log n}
\end{aligned}
$$

(recall $\kappa_0 = 40\kappa$, see (175)), and the similar lower bound holds. This gives (207).

Hence, (206) and (207) imply that for some $\iota'(\kappa)$ with $\lim_{\kappa\to 0}\iota' = 0$, under the event $\mathcal{A}_6$ (recall that $\mathcal{A}_6 = \mathcal{A}'\cap\mathcal{A}_3\cap\mathcal{A}_4\cap\mathcal{A}_5$) ,

$$
\sum_{i\neq j,i,j\in T}\Big(|Z_{ij}^{(1)}| - \frac{1}{h(\delta)}\sqrt{2(1+\delta')\log n}\Big)^2 \le \Big(\iota'(\kappa)+\frac{C}{h(\delta)}\Big)(1+\delta')\log n.
$$

By Cauchy–Schwarz inequality, using the fact that $|T| \le \bar{k}$, under the event $\mathcal{A}_6$,

$$
\sum_{i\neq j,i,j\in T}\Big||Z_{ij}^{(1)}| - \frac{1}{h(\delta)}\sqrt{2(1+\delta')\log n}\Big| \le Ch(\delta)\sqrt{\Big(\iota'(\kappa)+\frac{C}{h(\delta)}\Big)(1+\delta')\log n}. \tag{208}
$$

Note that under the event $\mathcal{A}_5$,

$$\Big|\sum_{i\neq j,i,j\in T}|Z_{ij}|-\sum_{i,j\in T}|Z_{ij}^{(1)}|\Big|\le\sum_{i,j\in T}|Z_{ij}^{(2)}|\le\bar{k}^2\sqrt{\varepsilon\log\log n}$$
$$\le Ch(\delta)^2\sqrt{\varepsilon\log\log n}. \tag{209}$$

Hence, by the above two inequalities, under the event $\mathcal{A}_6$, for sufficiently large n,

$$\sum_{i\neq j,i,j\in T}\Big||Z_{ij}|-\frac{1}{h(\delta)}\sqrt{2(1+\delta')\log n}\Big|\le Ch(\delta)\sqrt{\Big(\iota'(\kappa)+\frac{C}{h(\delta)}\Big)(1+\delta')\log n}. \tag{210}$$

By (120) and recalling $\varepsilon\le\frac{1}{\delta^4}$, for large enough δ, the above implies

$$\sum_{i\neq j,i,j\in T}\Big||Z_{ij}|-\frac{1}{h(\delta)}\sqrt{2(1+\delta)\log n}\Big|\le Ch(\delta)\sqrt{\Big(\iota'(\kappa)+\frac{C}{h(\delta)}\Big)(1+\delta)\log n}. \tag{211}$$

In fact, by the triangle inequality, the difference between LHS of (210) and (211) is bounded by

$$\frac{\sqrt{\varepsilon}(1+\delta)\sqrt{\log n}}{h(\delta)}h(\delta)^2\overset{(116)}{\le}Ch(\delta)\frac{\sqrt{\log n}}{\delta}\overset{(9)}{\le}Ch(\delta)\sqrt{\Big(\iota'(\kappa)+\frac{C}{h(\delta)}\Big)(1+\delta)\log n}.$$

Since $\kappa_0=40\kappa$, by (191), under the event $\mathcal{A}_6$, we have $|T|\ge(1-c\kappa^{1/4})\bar{k}$. In addition, since $h(\delta)\ge c\delta^{1/3}$, one can simplify the term $\iota'(\kappa)+\frac{C}{h(\delta)}$ to $\zeta(\kappa)$ with $\lim_{\kappa\to0}\zeta(\kappa)=0$ if δ is chosen large enough depending on κ. Dividing both sides by $h(\delta)^2$ and using (205) completes the proof.

Remark 8.1 Note that (211) gives a bound depending on both δ and κ and only on taking δ large enough depending on κ yields the theorem. Further, even though we provided sharp bounds for both ℓ_1 and ℓ_2 norms, in fact, a lower bound for the former and an upper bound for the latter suffices.

9 Lower tail large deviations

We end with the short argument establishing the large deviation probability of the lower tail, Theorem 1.8.

Proof of Theorem 1.8 The upper bound is an easy consequence of the inequality (29). In fact, by Lemma 4.1, 4.2 and (29),

$$\mathbb{P}(\lambda_1(Z)\le\sqrt{2(1-\delta)\log n})\le\mathbb{P}(\max Z_{ij}\le\sqrt{2(1-\delta)\log n})$$

$$\leq \mathbb{E}\left(\mathbb{P}(\max Z_{ij} \leq \sqrt{2(1-\delta)\log n} \mid X)\mathbb{1}_{E_0}\right) + \mathbb{P}(E_0^c)$$
$$\leq e^{-c'\frac{n^\delta}{\sqrt{\log n}}} + e^{-cn}.$$

We now prove a matching lower bound. Define an event $\mathcal{S}_\delta$, measurable with respect to X, by

$$\mathcal{S}_\delta := \left\{\lambda_1(X) \leq (1+\delta)\sqrt{\frac{\log n}{\log\log n}}\right\}.$$

Notice that $\mathbb{P}(S_\delta) \to 1$ by Lemma 5.2. Since $\lambda_1(Z^{(2)}) \leq \sqrt{\varepsilon \log\log n} \cdot \lambda_1(X)$, conditionally on X, under the event $\mathcal{S}_\delta$, it holds that

$$\lambda_1(Z^{(2)}) \leq \sqrt{\varepsilon}(1+\delta)\sqrt{\log n}.$$

Since $\lambda_1(Z) \leq \lambda_1(Z^{(1)}) + \lambda_1(Z^{(2)})$,

$$\mathbb{P}(\lambda_1(Z) \leq \sqrt{2(1-\delta)\log n}) \geq \mathbb{P}(\lambda_1(Z^{(1)}) \leq \sqrt{2(1-\delta'')\log n},\ \lambda_1(Z^{(2)}) \leq \sqrt{\varepsilon}(1+\delta)\sqrt{\log n}),$$

where $\delta'' > 0$ is defined by $\sqrt{2(1-\delta'')} = \sqrt{2(1-\delta)} - \sqrt{\varepsilon}(1+\delta)$. Recalling the definition of $\mathcal{F}_0 = \mathcal{D}_{4\delta'} \cap \mathcal{C}_{4\delta'} \cap \mathcal{E}_{4\delta'} \cap Few-cycles$ from (77), analogously we define $\mathcal{F}_3 := \mathcal{D}_{4\delta''} \cap \mathcal{C}_{4\delta''} \cap \mathcal{E}_{4\delta''} \cap Few-cycles \cap \mathcal{S}_\delta$, we have

$$\mathbb{P}(\lambda_1(Z) \leq \sqrt{2(1-\delta)\log n}) \geq \mathbb{E}\left[\mathbb{P}(\lambda_1(Z^{(1)}) \leq \sqrt{2(1-\delta'')\log n} \mid X, X^{(1)})\mathbb{1}_{\mathcal{F}_3}\right]. \tag{212}$$

Above we use that $\mathcal{F}_3$ is measurable with respect to the sigma algebra generated by $\{X^{(1)}, X\}$. We now estimate $\mathbb{P}(\lambda_1(Z^{(1)}) \leq \sqrt{2(1-\delta'')\log n} \mid X, X^{(1)})$ under the event $\mathcal{F}_3$ and finally we will use that $\mathcal{F}_3$ is likely. We will crucially use throughout the proof that given $X^{(1)}$, $Z^{(1)}$ and X are conditionally independent.

Let $C_1, \cdots, C_m$ be $X^{(1)}$'s connected components and denote by k_i the size of maximal clique in C_i. Let

$$I := \{i = 1, \cdots, m : k_i \geq 3\}, \quad J := \{i = 1, \cdots, m : k_i = 2\}$$

and define $\xi := (2\varepsilon^{1/2} + 8\varepsilon\delta'')^{1/4}$. By Proposition 5.7 with $\gamma = \varepsilon$ and $\eta = \varepsilon^{1/4}$, for sufficiently small $\varepsilon > 0$, under the event $\mathcal{F}_3$, for $i \in I$,

$$\mathbb{P}(\lambda_1(Z_i^{(1)}) \geq \sqrt{2(1-\delta'')\log n} \mid X, X^{(1)}) < n^{-\frac{1}{2}(1-\xi)^2(1-\delta'') + \frac{1+4\delta''}{2}\varepsilon^{1/2} + \varepsilon} \tag{213}$$

using the fact that $\frac{k}{k-1} \geq 1$, and for $i \in J$,

$$\mathbb{P}(\lambda_1(Z_i^{(1)}) \geq \sqrt{2(1-\delta'')\log n} \mid X, X^{(1)}) < n^{-(1-\xi)^2(1-\delta'') + \frac{1+4\delta''}{2}\varepsilon^{1/2} + \varepsilon}. \tag{214}$$

Since $|I| < \log n$ under the event $Few - Cycles$, by (213) and (214),

$$\begin{aligned}\mathbb{P}(\lambda_1(Z_i^{(1)}) &\le \sqrt{2(1-\delta)\log n},\ \forall i \mid X, X^{(1)}) \\ &> (1 - n^{-(1-\xi)^2(1-\delta'')+\frac{1+4\delta''}{2}\varepsilon^{1/2}+\varepsilon})^n (1 - n^{-\frac{1}{2}(1-\xi)^2(1-\delta'')+\frac{1+4\delta''}{2}\varepsilon^{1/2}+\varepsilon})^{\log n} \\ &\ge \frac{1}{2}\exp(-n^{1-(1-\xi)^2(1-\delta'')+\frac{1+4\delta''}{2}\varepsilon^{1/2}+\varepsilon}). \end{aligned} \tag{215}$$

Since $\mathbb{P}(\mathcal{F}_3) \ge \frac{1}{2}$ and $\varepsilon > 0$ is arbitrary small, by (212) and (215), proof is concluded. □

Acknowledgements The authors thank an anonymous referee for several useful comments that helped improve the article. They also thank Noga Alon for pointing out the classical reference [37]. SG is partially supported by NSF grant DMS-1855688, NSF CAREER Award DMS-1945172 and a Sloan Research Fellowship. KN is supported by the National Research Foundation of Korea (NRF-2019R1A6A1A10073887, NRF-2019R1A5A1028324). This work was initiated when SG was participating in the Probability, Geometry, and Computation in High Dimensions program at the Simons Institute in Fall 2020.

Appendix A. Key estimates

In this appendix, we include the outstanding proofs of basic properties about Gaussian random variables. as well as the proof of Lemma 4.2 involving a straightforward application of Chernoff's bound.

Proof of Lemma 4.1 Recalling the basic tail bounds from (24), for some constant $c_1 > 0$,

$$\begin{aligned}\mathbb{P}(\max_{i=1,\cdots,m} X_i \ge \sqrt{2(1+\delta)\log n}) &= 1 - (1 - \mathbb{P}(X_1 \ge \sqrt{2(1+\delta)\log n}))^m \\ &\ge c_1 \frac{1}{n^\delta \sqrt{\log n}}.\end{aligned}$$

Similarly, for some constant $c_2 > 0$,

$$\mathbb{P}(\max_{i=1,\cdots,m} X_i \le \sqrt{2(1-\delta)\log n}) = (1 - \mathbb{P}(X_1 \ge \sqrt{2(1-\delta)\log n}))^m \le e^{-c_2 \frac{n^\delta}{\sqrt{\log n}}}.$$

□

Proof of Lemma 4.2 We use the Chernoff's bound for Bernoulli variables for $q > p$:

$$\mathbb{P}(\mathrm{Bin}(m, p) \ge mq) \le e^{-mI_p(q)}, \tag{216}$$

where $I_p(x) := x \log \frac{x}{p} + (1-x)\log \frac{1-x}{1-p}$ is the relative entropy function. Thus,

$$\mathbb{P}\Big(\mathrm{Bin}\Big(\frac{n(n-1)}{2}, 1 - \frac{d}{n}\Big) \ge \frac{n(n-1)}{2}\Big(1 - \frac{d}{4n}\Big)\Big) \le e^{-\frac{n(n-1)}{2} I_{1-\frac{d}{n}}(1-\frac{d}{4n})}, \tag{217}$$

Using $\log(1+x) \geq \frac{x}{2}$ for small positive x,

$$I_{1-\frac{d}{n}}\Big(1-\frac{d}{4n}\Big) \geq \Big(1-\frac{d}{4n}\Big)\frac{3d}{8(n-d)} + \frac{d}{4n}\log\frac{1}{4} \geq \frac{C_1}{n-d} - \frac{C_2}{n^2}. \tag{218}$$

Hence, by (217) and (218), there exists a constant $c > 0$ such that for sufficiently large n,

$$\mathbb{P}\Big(\mathrm{Bin}\Big(\frac{n(n-1)}{2}, 1-\frac{d}{n}\Big) \geq \frac{n(n-1)}{2}\Big(1-\frac{d}{4n}\Big)\Big) \leq e^{-cn}.$$

This implies that

$$\mathbb{P}\Big(\mathrm{Bin}\Big(\frac{n(n-1)}{2}, \frac{d}{n}\Big) \leq \frac{n(n-1)}{2}\frac{d}{4n}\Big) \leq e^{-cn},$$

which concludes the proof. □

Proof of Lemma 5.1 Recall that we are aiming to show

$$\mathbb{P}(\tilde{Y}_1^2 + \cdots + \tilde{Y}_m^2 \geq L) \leq C^m e^{-\frac{1}{2}L} e^{\frac{1}{2}m}\Big(\frac{L}{m}\Big)^m e^{\frac{1}{2}\varepsilon m \log\log n},$$

and in particular, for any $a, b, c > 0$, if $m \leq b\frac{\log n}{\log\log n} + c$ and $L = a\log n$, then, for any $\gamma > 0$, for sufficiently large n,

$$\mathbb{P}(\tilde{Y}_1^2 + \cdots + \tilde{Y}_m^2 \geq a\log n) \leq n^{-\frac{a}{2}+\frac{\varepsilon b}{2}+\gamma}. \tag{219}$$

By exponential Chebyshev's bound, for any $t > 0$,

$$\mathbb{P}(\tilde{Y}_1^2 + \cdots + \tilde{Y}_m^2 \geq L) \leq e^{-tL}(\mathbb{E}e^{t\tilde{Y}_1^2})^m. \tag{220}$$

Using the lower bound for the tail (24), the probability density function of $\tilde{Y}$, denoted by $\tilde{f}(x)$ for $|x| \geq \sqrt{\varepsilon\log\log n}$, satisfies

$$\tilde{f}(x) \leq \frac{C}{(\sqrt{\varepsilon\log\log n})^{-1}e^{-\frac{1}{2}\varepsilon\log\log n}}e^{-\frac{1}{2}x^2} = C\sqrt{\varepsilon\log\log n}\,e^{\frac{1}{2}\varepsilon\log\log n}e^{-\frac{1}{2}x^2}.$$

Hence, using the upper bound for the tail (24), by making a change of variable $x = \frac{1}{\sqrt{1-2t}}y$,

$$\begin{aligned}\mathbb{E}e^{t\tilde{Y}_1^2} &\leq C\sqrt{\varepsilon\log\log n}\,e^{\frac{1}{2}\varepsilon\log\log n}\int_{\sqrt{\varepsilon\log\log n}}^{\infty} e^{tx^2}e^{-\frac{1}{2}x^2}dx \\ &= C\sqrt{\varepsilon\log\log n}\,e^{\frac{1}{2}\varepsilon\log\log n}\frac{1}{\sqrt{1-2t}}\int_{\sqrt{1-2t}\sqrt{\varepsilon\log\log n}}^{\infty} e^{-\frac{1}{2}y^2}dy\end{aligned}$$

$$\leq C\sqrt{\varepsilon \log\log n}\, e^{\frac{1}{2}\varepsilon \log\log n} \frac{1}{\sqrt{1-2t}} \frac{1}{\sqrt{1-2t}\sqrt{\varepsilon \log\log n}} e^{-\frac{1}{2}(1-2t)\varepsilon \log\log n}$$
$$= C\frac{1}{1-2t} e^{t\varepsilon \log\log n}.$$

Applying this to (220),

$$\mathbb{P}(\tilde{Y}_1^2 + \cdots + \tilde{Y}_m^2 \geq L) \leq C^m e^{-tL} \frac{1}{(1-2t)^m} e^{t\varepsilon m \log\log n}.$$

We take $t = \frac{1}{2}(1 - \frac{m}{L}) < \frac{1}{2}$ (recall that $L > m$) in order to balance two terms e^{-tL} and $\frac{1}{(1-2t)^m}$. We conclude the proof of (38).

We now show (219). We first check that for any $L > 0$, a function $x \mapsto (\frac{L}{x})^x$ is increasing on $(0, \frac{L}{e})$. This is because the derivative of $x \log(\frac{L}{x})$, which is given by $\log(\frac{L}{x}) - 1$, is positive for $x \in (0, \frac{L}{e})$. Hence, for any $\gamma > 0$, for sufficiently large n, the LHS of (219) is bounded by

$$C^{b\frac{\log n}{\log\log n}+c} n^{-\frac{a}{2}+\frac{\varepsilon b}{2}} n^{\frac{b}{2\log\log n}} \left(\frac{a}{b}\log\log n\right)^{b\frac{\log n}{\log\log n}+c} \leq n^{-\frac{a}{2}+\frac{\varepsilon b}{2}+\gamma}.$$

Here, we used the fact that for large n, $(c_1 \log\log n)^{c_2 \frac{\log n}{\log\log n}} \leq n^{\frac{\gamma}{2}}$. □

References

1. Alt, J., Ducatez, R., Knowles, A.: Extremal eigenvalues of critical erdős-rényi graphs. arXiv preprint arXiv:1905.03243, (2019)
2. Ben Arous, G., Dembo, A., Guionnet, A.: Aging of spherical spin glasses. Probab. Theory Relat. Fields **120**(1), 1–67 (2001)
3. Ben Arous, G., Guionnet, A.: Large deviations for Wigner's law and Voiculescu's non-commutative entropy. Probab. Theory Relat. Fields **108**(4), 517–542 (1997)
4. Augeri, Fanny: Large deviations principle for the largest eigenvalue of wigner matrices without gaussian tails. Electron. J. Probab. **21**, 49 (2016)
5. Augeri, F.: Nonlinear large deviation bounds with applications to traces of wigner matrices and cycles counts in Erdős-Rényi graphs. Ann. Probab. (2020)
6. Augeri, F., Guionnet, A., Husson, J.: Large deviations for the largest eigenvalue of sub-gaussian matrices. arXiv preprint arXiv:1911.10591 (2019)
7. Austin, T.: The structure of low-complexity gibbs measures on product spaces. Ann. Probab. **47**(6), 4002–4023 (2019)
8. Bandeira, A.S., Van Handel, R.: Sharp nonasymptotic bounds on the norm of random matrices with independent entries. Ann. Probab. **44**(4), 2479–2506 (2016)
9. Basak, A., Basu, R.: Upper tail large deviations of the cycle counts in Erdős-Rényi graphs in the full localized regime. arXiv:1912.11410 (2019)
10. Basak, A., Mukherjee, S.: Universality of the mean-field for the potts model. Probab. Theory Relat. Fields **168**(3–4), 557–600 (2017)
11. Benaych-Georges, F., Bordenave, C., Knowles, A.: Largest eigenvalues of sparse inhomogeneous erdős-rényi graphs. Ann. Probab. **47**(3), 1653–1676 (2019)
12. Benaych-Georges, F., Bordenave, C., Knowles, A.: Spectral radii of sparse random matrices. In Annales de l'Institut Henri Poincaré, Probabilités et Statistiques, vol. 56, pp. 2141–2161. Institut Henri Poincaré (2020)

13. Bhattacharya, B.B, Bhattacharya, S., Ganguly, S.: Spectral edge in sparse random graphs: Upper and lower tail large deviations. Ann. Probab. **49**(4), 1847–1885 (2021)
14. Bhattacharya, B.B., Ganguly, S.: Upper tails for edge eigenvalues of random graphs. SIAM J. Discret. Math. (to appear) (2020)
15. Bhattacharya, B.B., Ganguly, S., Lubetzky, E., Zhao, Y.: Upper tails and independence polynomials in random graphs. Adv. Math. **319**, 313–347 (2017)
16. Bollobás, B.: Random Graphs, vol. 73. Cambridge University Press, Cambridge (2001)
17. Bordenave, C., Caputo, P.: A large deviation principle for Wigner matrices without gaussian tails. Ann. Probab. **42**(6), 2454–2496 (2014)
18. Bordenave, C., Caputo, P.: Large deviations of empirical neighborhood distribution in sparse random graphs. Probab. Theory Relat. Fields **163**(1–2), 149–222 (2015)
19. Bordenave, C., Sen, A., Virág, B.: Mean quantum percolation. J. Eur. Math. Soc. **19**(12), 3679–3707 (2017)
20. Chatterjee, S.: Superconcentration and Related Topics, vol. 15. Springer, New York (2014)
21. Chatterjee, S., Dembo, A.: Nonlinear large deviations. Adv. Math. **299**, 396–450 (2016)
22. Chatterjee, S., Varadhan, S.R.S.: The large deviation principle for the Erdős-Rényi random graph. Eur. J. Combin. **32**(7), 1000–1017 (2011)
23. Chatterjee, S., Varadhan, S.R.S.: Large deviations for random matrices. Commun. Stoch. Anal. **6**(1), 1–13 (2012)
24. Cook, N., Dembo, A.: Large deviations of subgraph counts for sparse Erdős-Rényi graphs. arXiv:1809.11148 (2018)
25. Dembo, A., Zeitouni, O.: Large Deviations Techniques and Applications, volume 38 of Stochastic Modelling and Applied Probability. Springer, Berlin (2010). Corrected reprint of the second (1998) edition
26. Eldan, R.: Gaussian-width gradient complexity, reverse log-Sobolev inequalities and nonlinear large deviations. Geom. Funct. Anal. (to appear) (2018)
27. Erdős, L., Knowles, A., Yau, H.-T., Yin, J.: Spectral statistics of erdős-rényi graphs II: eigenvalue spacing and the extreme eigenvalues. Commun. Math. Phys. **314**(3), 587–640 (2012)
28. Erdős, L., Knowles, A., Yau, H.-T., Yin, J.: Spectral statistics of erdős-rényi graphs I: local semicircle law. Ann. Probab. **41**(3B), 2279–2375 (2013)
29. Frieze, A., Karoński, M.: Introduction to Random Graphs. Cambridge University Press, Cambridge (2016)
30. Guionnet, A., Husson, J.: Large deviations for the largest eigenvalue of rademacher matrices. Ann. Probab. (to appear) (2020)
31. Harel, M., Mousset, F., Samotij, W.: Upper tails via high moments and entropic stability. arXiv:1904.08212 (2019)
32. Krivelevich, M., Sudakov, B.: The largest eigenvalue of sparse random graphs. Comb. Probab. Comput. **12**(1), 61–72 (2003)
33. Latała, R.: Some estimates of norms of random matrices. Proc. Am. Math. Soc. **133**(5), 1273–1282 (2005)
34. Latała, R., van Handel, R., Youssef, P.: The dimension-free structure of nonhomogeneous random matrices. Invent. Math. **214**(3), 1031–1080 (2018)
35. Lubetzky, E., Zhao, Y.: On replica symmetry of large deviations in random graphs. Rand. Struct. Algorith. **47**(1), 109–146 (2015)
36. Lubetzky, E., Zhao, Y.: On the variational problem for upper tails in sparse random graphs. Rand. Struct. Algorith. **50**(3), 420–436 (2017)
37. Motzkin, T.S., Straus, E.G.: Maxima for graphs and a new proof of a theorem of Turán. Can. J. Math. **17**, 533–540 (1965)
38. Reimer, D.: Proof of the van den berg-kesten conjecture. Comb. Probab. Comput. **9**(1), 27–32 (2000)
39. Seginer, Y.: The expected norm of random matrices. Comb. Probab. Comput. **9**(2), 149–166 (2000)
40. Tikhomirov, K., Youssef, P.: Outliers in spectrum of sparse wigner matrices. Rand. Struct. Algorith. (2020)

41. Van Handel, R.: On the spectral norm of gaussian random matrices. Trans. Am. Math. Soc. **369**(11), 8161–8178 (2017)
42. Yan, J.: Nonlinear large deviations: Beyond the hypercube. Ann. Appl. Probab. (to appear) (2020)