Bernoulli **19**(1), 2013, 275–294
DOI: 10.3150/11-BEJ394

On the maximal size of large-average and ANOVA-fit submatrices in a Gaussian random matrix

XING SUN[1] and ANDREW B. NOBEL[2]

[1] *Merck & Co., Inc., One Merck Drive, Whitehouse Station, NJ 08889, USA. E-mail: xing_sun@merk.com*
[2] *Department of Statistics and Operations Research, University of North Carolina, Chapel Hill, NC 27599-3260, USA. E-mail: nobel@email.unc.edu*

We investigate the maximal size of distinguished submatrices of a Gaussian random matrix. Of interest are submatrices whose entries have an average greater than or equal to a positive constant, and submatrices whose entries are well fit by a two-way ANOVA model. We identify size thresholds and associated (asymptotic) probability bounds for both large-average and ANOVA-fit submatrices. Probability bounds are obtained when the matrix and submatrices of interest are square and, in rectangular cases, when the matrix and submatrices of interest have fixed aspect ratios. Our principal result is an almost sure interval concentration result for the size of large average submatrices in the square case.

Keywords: analysis of variance; data mining; Gaussian random matrix; large average submatrix; random matrix theory; second moment method

1. Introduction

A Gaussian random matrix (GRM) is a matrix whose elements are i.i.d. standard normal random variables. Gaussian random matrices have been a fixture in the application and theory of multivariate analysis for many years. Recent work in the field of random matrix theory has provided a wealth of information about the eigenvalues and eigenvectors of Gaussian and more general, random matrices (see, e.g., Anderson, Guionnet and Zeitouni [2]). This paper considers a different problem, namely, the maximal size of distinguished submatrices in a Gaussian random matrix. We consider submatrices that are distinguished in one of two ways: (i) the average of their entries is greater than or equal to a positive constant, or (ii) the optimal two-way ANOVA fit of their entries has average squared residual less than a positive constant.

Our goal is to identify maximal size thresholds, and associated probability bounds, for large average and ANOVA-fit submatrices. Results are obtained when the matrix and the submatrices of interest are square, and when the matrix and the submatrices of interest are rectangular with fixed aspect ratios. In each case, the maximal size of a distinguished submatrix grows logarithmically with the dimension of the matrix and depends, in a polynomial fashion, on the inverse of the constant that constitutes the distinguishability threshold. In the rectangular case, the aspect ratio of the submatrix plays a more critical role than the aspect ratio of the matrix itself. Our principal result establishes almost sure upper and lower bounds for the size of large average submatrices in the square case. In particular, for $n \times n$ Gaussian random matrices, we establish that the size of

1350-7265 © 2013 ISI/BS

the largest square submatrix with average greater than a positive constant τ is eventually almost surely within an interval of fixed width that contains the critical value $4\tau^{-2}(\ln n - \ln(4\tau^{-2}\ln n))$.

Results of the sort established here fall outside the purview of random matrix theory and its techniques. Nevertheless, random matrix theory does provide some insight into the logarithmic scale of large average submatrices. This is discussed briefly in Section 1.2 below.

1.1. Bipartite graphs

Our results on large average submatrices can also be expressed in graph-theoretic terms, as every $m \times n$ matrix X is associated in a natural way with a bipartite graph $G = (V, E)$. In particular, the vertex set V of G is the disjoint union of two sets V_1 and V_2, with $|V_1| = m$ and $|V_2| = n$, corresponding to the rows and columns of X, respectively. For each row $i \in V_1$ and column $j \in V_2$ there is an edge $(i, j) \in E$ with weight $x_{i,j}$. There are no edges between vertices in V_1 or between vertices in V_2. With this association, large average submatrices of X are in 1: 1 correspondence with subgraphs of G having large average edge-weight. The complexity of finding the largest subgraph of G whose average edge weight is greater than a threshold appears to be unknown. However, it is shown in [4] that a slight variation of this problem, namely finding the maximum edge weight subgraph in a general bipartite matrix, is NP-complete. A randomized, polynomial time algorithm that finds a subgraph whose edge weight is within a constant factor of the optimum is described in [1], but this algorithm cannot readily be adapted to the problem of identifying the large average submatrices considered here.

1.2. Size thresholds and random matrix theory

The results of this paper are combinatorial in nature and do not rely on the spectral techniques employed in random matrix theory. Nevertheless, existing results in random matrix theory provide insights into the relationship between the large average submatrices studied here and the singular value decomposition. These results indicate that there is a significant gap between the logarithmic size thresholds at which large average submatrices become significant, and the root-n size thresholds at which they are detectable by standard spectral methods.

Let W be an $m \times n$ Gaussian random matrix, and let $\tau > 0$ be fixed. Define a rank-one matrix $S = (1 + \delta)\tau ab^t$, where $\delta > 0$ and $a \in \{0, 1\}^m$, $b \in \{0, 1\}^n$ are indicator vectors having k and l non-zero components, respectively. The outer product ab^t defines a submatrix U whose rows and columns are indexed by the indicator vectors a and b. Let $Y = W + S$ be the sum of W and S, which we regard as a perturbed version of S. Suppose that the dimensions m, n grow in such a way that $m/n \to \alpha$ with $\alpha \in [1, \infty)$, and that the dimensions k, l grow in such a way that $k/\log n \to \infty$ and k/l remains bounded away from zero and infinity. It follows from Proposition 4 that the probability of finding any $k \times l$ submatrix with average greater than τ in the unperturbed matrix W is vanishingly small. On the other hand, the average of the $k \times l$ submatrix U of Y has distribution $\mathcal{N}((1 + \delta)\tau, (kl)^{-1})$, which is greater than τ with probability very close to one when k and l are large. We might expect to see evidence of the submatrix U in the first singular value of the matrix Y, or its associated left and right singular vectors. However, in a sense we make precise below, this is not the case.

Let $s_1(V) \geq \cdots \geq s_m(V)$ denote the ordered singular values of an $m \times n$ matrix V, and let $\|V\|_F = \sum_{i,j} v_{i,j}^2$ denote its Frobenius norm. As $s_1(\cdot)$ is a norm, we have

$$|s_1(Y) - s_1(W)| \leq s_1(Y - W) \leq \|Y - W\|_F = (1+\delta)^2 \tau^2 kl, \tag{1}$$

where the second inequality makes use of the fact that the Frobenius norm of a matrix is the sum of the squares of its singular values. By a basic result of Geman [5],

$$\frac{s_1(W)}{n^{1/2}} \to (1 + \alpha^{1/2}) \tag{2}$$

with probability one as n tends to infinity. If $k = \mathrm{o}(m^{1/2})$ and $l = \mathrm{o}(n^{1/2})$, inequality (1) implies that $n^{-1/2}|s_1(Y) - s_1(W)| \to 0$ with probability one, and therefore (2) holds with Y in place of W. In other words, for fixed τ, and dimensions k, l such that $\log n \ll k, l \ll n^{1/2}$, the embedded submatrix U of Y is highly significant, but has no effect on the limiting behavior of $s_1(Y)$. Under the same conditions, U is also not recoverable from the top singular vectors of Y. To be precise, let u_1 and v_1 be the left and right singular vectors of Y corresponding to the maximum singular value $s_1(Y)$. Using results of Paul [8] on the singular vectors of spiked population models, it can be shown that $a^t u_1$ and $b^t v_1$ tend to zero in probability as n tends to infinity. Thus the row and column index vectors of U are asymptotically orthogonal to the first left and right singular vectors of Y.

1.3. Overview

The next section contains probability bounds and a finite interval concentration result for the size of large average submatrices in the square case. Size thresholds and probability bounds for ANOVA submatrices in the square case are presented in Section 3. Thresholds and bounds in the rectangular case are given in Section 4. Sections 5–7 contain the proofs of the main results.

2. Thresholds and bounds for large average submatrices

Let $W = \{w_{i,j}\colon i, j \geq 1\}$ be an infinite array of independent $\mathcal{N}(0,1)$ random variables, and for $n \geq 1$, let $W_n = \{w_{i,j}\colon 1 \leq i, j \leq n\}$ be the $n \times n$ Gaussian random matrix equal to upper left-hand corner of W. (The almost sure asymptotics of Theorem 1 requires consideration of matrices W_n that are derived from a fixed, infinite array.) A submatrix of W_n is an indexed collection $U = \{w_{i,j}\colon i \in A, j \in B\}$ where $A, B \subseteq \{1, \ldots, n\}$. The Cartesian product $C = A \times B$ will be called the index set of U, and we will write $U = W_n[C]$. The dimension of U is $|A| \times |B|$, where $|A|, |B|$ denote the cardinality of A and B, respectively. Note that rows A need not be contiguous, and that the same is true of columns B.

Definition. *For any submatrix U of W_n with index set $C = A \times B$, let*

$$F(U) = \frac{1}{|C|} \sum_{(i,j) \in C} w_{i,j} = \frac{1}{|A||B|} \sum_{i \in A, j \in B} w_{i,j}$$

be the average of the entries of U*. Note that* $F(U) \sim \mathcal{N}(0, |C|^{-1})$.

We are interested in the maximal size of square submatrices whose averages exceed a fixed threshold. This motivates the following definition.

Definition. *Fix* $\tau > 0$ *and* $n \geq 1$*. Let* $K_\tau(W_n)$ *be the largest* $k \geq 0$ *such that* W_n *contains a* $k \times k$ *submatrix* U *with* $F(U) \geq \tau$.

As the rows and columns of a submatrix need not be contiguous, the statistic $K_\tau(W_n)$ is invariant under row and column permutations of W_n. Our immediate goal is to obtain bounds on the probability that $K_\tau(W_n)$ exceeds a given threshold and to identify a threshold for $K_\tau(W_n)$ that governs its asymptotic behavior. To this end, we begin the analysis of $K_\tau(W_n)$ using standard first moment-type arguments, which are detailed below.

Let $\Gamma_k(n, \tau)$ be the number of $k \times k$ submatrices in W_n having an average greater than or equal to τ. We begin by identifying the value of k for which $E\Gamma_k(n, \tau)$ is approximately equal to one. Let $\mathcal{S}_k$ denote the set of all $k \times k$ submatrices of W_n. Then

$$\Gamma_k(n, \tau) = \sum_{U \in \mathcal{S}_k} I\{F(U) \geq \tau\}, \tag{3}$$

and, consequently,

$$E\Gamma_k(n, \tau) = |\mathcal{S}_k| \cdot P\bigl(F(W_k) \geq \tau\bigr) = \binom{n}{k}^2 \bigl(1 - \Phi(\tau k)\bigr) \leq \binom{n}{k}^2 \mathrm{e}^{-\tau^2 k^2/2}, \tag{4}$$

where in the last step we have used a standard bound on $1 - \Phi(\cdot)$. For $s \in (0, n)$, define

$$\phi_{n,\tau}(s) = (2\pi)^{-1/2} n^{n+1/2} s^{-s-1/2} (n-s)^{-(n-s)-1/2} \mathrm{e}^{-\tau^2 s^2/4}. \tag{5}$$

Using the Stirling approximation of $\binom{n}{k}$, it is easy to see that $\phi_{n,\tau}(k)$ is an approximation of the square root of the final expression in (4). In particular, the rightmost expression in (4) is less than $2\phi_{n,\tau}(k)^2$. With this in mind, let $s(n, \tau)$ be any positive, real root of the equation

$$\phi_{n,\tau}(s) = 1. \tag{6}$$

The next result shows that $s(n, \tau)$ exists and is unique, and it provides an explicit expression for its value when τ is fixed and n is large.

Lemma 1. *Let* $\tau > 0$ *be fixed. When* n *is sufficiently large, equation* (6) *has a unique root* $s(n, \tau)$*, and*

$$s(n, \tau) = \frac{4}{\tau^2} \ln n - \frac{4}{\tau^2} \ln\left(\frac{4}{\tau^2} \ln n\right) + \frac{4}{\tau^2} + \mathrm{o}(1), \tag{7}$$

where $\mathrm{o}(1) \to 0$ *as* $n \to \infty$.

We show below that the asymptotic behavior of the random variables $K_\tau(W_n)$ is governed by the root $s(n,\tau)$ of equation (6). To begin, note that for values of k greater than $s(n,\tau)$, the expected number of $k \times k$ submatrices U of W_n with $F(U) \geq \tau$ is less than one. The next proposition shows that the probability of seeing such large submatrices is small.

Proposition 1. *Let $\tau > 0$ be fixed. When n is sufficiently large,*

$$P\big(K_\tau(W_n) \geq s(n,\tau) + r\big) \leq 2\mathrm{e}^{2/\tau^2} n^{-2r} \left(\frac{4 \ln n}{\tau^2}\right)^{2r}$$

for every $r = 1, \dots, n$.

The proofs of Lemma 1 and Proposition 1 are given in Section 5. The arguments refine those in [11], with adaptations to the present setting. The asymptotic nature of the bound in Proposition 1 results from the o(1) term in $s(n,\tau)$, and, in particular, approximations arising from the general form of Stirling's formula. Using a more elementary bound, one may readily obtain a non-asymptotic result for a size threshold that includes only the leading term of $s(n,\tau)$.

Proposition 2. *Let $\tau > 0$ be fixed. Then*

$$P\left(K_\tau(W_n) \geq \frac{4}{\tau^2} \ln n + r\right) \leq n^{-2r}$$

for every $n, r \geq$ such that $\frac{4}{\tau^2} \ln n + r > 2$.

It follows from Proposition 1 and the Borel–Cantelli lemma that, with probability one, $K_\tau(W_n)$ is eventually less than or equal to $\lceil s(n,\tau) \rceil + 1 \leq s(n,\tau) + 2$. With this bound in mind, it is of interest to know more about the asymptotic behavior of $K_\tau(W_n)$. It turns out that the limiting distribution of $K_\tau(W_n)$ is essentially degenerate. Our principal result, stated in Theorem 1 below, makes use of a second moment argument in order to obtain an almost sure lower bound on $K_\tau(W_n)$ that is within a constant factor of the upper bound derived from Proposition 1. The proof is given in Section 7.

Theorem 1. *Let W_n, $n \geq 1$, be Gaussian random matrices derived from an infinite array W, and let $\tau > 0$ be fixed. With probability one, when n is sufficiently large,*

$$s(n,\tau) - \frac{4}{\tau^2} - \frac{12 \ln 2}{\tau^2} - 4 \leq K_\tau(W_n) \leq s(n,\tau) + 2. \tag{8}$$

Note that the difference between the upper and lower bounds in Theorem 1 is a constant that depends on τ, but is *independent* of the matrix dimension n. In particular, the values of the random variable $K_\tau(W_n)$ are eventually concentrated on an interval that contains $s(n,\tau)$ and whose width is independent of n. It follows from Theorem 1 that

$$\frac{K_\tau(W_n)}{4\tau^{-2} \log n} \to 1$$

almost surely as $n \to \infty$. The lower bound in Theorem 1 can be slightly improved. An examination of the argument in Lemma 4 in Section 7 shows the inequality of the theorem still holds if the quantity $12 \ln 2$ is replaced with any constant greater than $8 \ln 2$.

Extending earlier work of Dawande *et al.* [4] and Koyuturk *et al.* [6], Sun and Nobel [10,11] obtained a similar, two-point concentration result for the size of largest square submatrix of ones in an i.i.d. Bernoulli random matrix. Bollobás and Erdős [3] and Matula [7], established analogous results for the clique number of a regular random graph; see [11] for additional references to work in the binary case. The proof of Theorem 1 relies on a second moment argument, but differs from the proofs of these earlier results due to the continuous setting. In particular, the proof makes use of the fact that, under the Gaussian assumption made here, for any $k \times k$ submatrix U of W, there exists an upper bound and a lower bound on $P(F(U) \geq \tau)$ whose ratio is of order τk.

3. Thresholds and bounds for ANOVA submatrices

In this section, we derive bounds like those in Proposition 1 for the size of submatrices whose entries are well fit by a two-way analysis of variance (ANOVA) model. A statistical introduction to ANOVA can be found in Scheffé [9]. Roughly speaking, the ANOVA criterion identifies submatrices whose rows (and columns) are shifts of each other.

Definition. *For a submatrix U of W_n with index set $A \times B$, define*

$$G(U) = \min\left\{\frac{1}{(|A|-1)(|B|-1)} \sum_{i \in A, j \in B} (w_{ij} - a_i - b_j - c)^2\right\},$$

where the minimum is taken over all real constants $\{a_i\colon\ i \in A\}$, $\{b_j\colon\ j \in B\}$ and c.

Under the ANOVA criterion, a submatrix U will warrant interest if $G(U)$ is less than a predefined threshold. Note that by standard arguments,

$$G(U) = \frac{1}{(|A|-1)(|B|-1)} \sum_{i \in A, j \in B} (w_{ij} - \overline{w}_{i\cdot} - \overline{w}_{\cdot j} + \overline{w}_{\cdot\cdot})^2,$$

where $\overline{w}_{i\cdot}$, $\overline{w}_{\cdot j}$ and $\overline{w}_{\cdot\cdot}$ denote the row, column, and the full submatrix averages, respectively.

Definition. *Given $0 < \tau < 1$, let $L_\tau(W_n)$ be the largest value of k such that W_n contains a $k \times k$ submatrix U with $G(U) \leq \tau$.*

Arguments similar to those in the proof of Proposition 1, in conjunction with a probability upper bound on the left tail of a χ^2 distribution, establish the following bound on $L_\tau(W_n)$. The proof is given in Section 6.

Proposition 3. *Let $\tau > 0$ be fixed. When n is sufficiently large,*

$$P\big(L_\tau(W_n) \geq t(n,\tau) + r\big) \leq 2\mathrm{e}^{1+2/h(\tau)} \left(\frac{\ln n}{h(\tau)}\right)^{2r+2} n^{-2r} \tag{9}$$

for every $r = 1, \ldots, n$, where

$$t(n,\tau) = \frac{4}{h(\tau)} \ln n - \frac{4}{h(\tau)} \ln\left(\frac{4}{h(\tau)} \ln n\right) + \frac{4}{h(\tau)} + 2$$

and

$$h(\tau) = 1 - \tau - \log(2 - \tau). \tag{10}$$

Proposition 3 and the Borel–Cantelli lemma imply that $L_\tau(W_n) \leq t(n,\tau) + 1$, eventually, almost surely. The arguments used to lower bound $K_\tau(W_n)$ in Theorem 1 do not extend readily to $L_\tau(W_n)$; we are not aware if a similar interval-concentration result holds in this case.

4. Thresholds and bounds for rectangular submatrices

The probability bounds of Propositions 1 and 3 can be extended to non-square submatrices of non-square matrices by adapting the methods of proof, detailed in Sections 5 and 6, respectively. We present the resulting bounds below, without proof. Similar results concerning submatrices of 1s in binary matrices can be found in [11].

Definition. *Let $W(m,n)$ denote an $m \times n$ Gaussian random matrix, and let $\alpha > 0$ and $\beta \geq 1$ be fixed aspect ratios for the sample matrix and target submatrix, respectively.*

a. *For $\tau > 0$, let $K_\tau(W\colon n,\alpha,\beta)$ be the largest integer k such that there exists a $\lceil \beta k \rceil \times k$ submatrix U in $W(\lceil \alpha n \rceil, n)$ with $F(U) \geq \tau$.*

b. *For $0 < \tau < 1$, let $L_\tau(W\colon n,\alpha,\beta)$ be the largest integer k such that there exists a $\lceil \beta k \rceil \times k$ submatrix U in $W(\lceil \alpha n \rceil, n)$ with $G(U) \leq \tau$.*

Proposition 4. *Fix $\tau > 0$ and any $\varepsilon > 0$. When n is sufficiently large,*

$$P\big(K_\tau(W\colon n,\alpha,\beta) \geq s(n,\tau,\alpha,\beta) + r\big) \leq n^{-(\beta+1)r} \left(\frac{\ln n}{\tau^2}\right)^{(\beta+1+\varepsilon)r}$$

for each $1 \leq r \leq n$, where

$$s(n,\tau,\alpha,\beta) = \frac{2(1+\beta^{-1})}{\tau^2} \ln n - \frac{2(1+\beta^{-1})}{\tau^2} \ln\left[\frac{2(1+\beta^{-1})}{\tau^2} \ln n\right] + \frac{2}{\tau^2} \ln \alpha + C_1(\beta,\tau)$$

for some constant $C_1(\beta,\tau) > 0$.

Proposition 5. *Fix $0 < \tau < 1$ and any $\varepsilon > 0$. When n is sufficiently large,*

$$P\big(L_\tau(W\colon n, \alpha, \beta) \geq t(n, \tau, \alpha, \beta) + r\big) \leq n^{-(\beta+1)r} \left(\frac{\ln n}{h(\tau)}\right)^{(\beta+1+\varepsilon)r}$$

for each $1 \leq r \leq n$, where

$$t(n, \tau, \alpha, \beta) = \frac{2(1+\beta^{-1})}{h(\tau)} \ln n - \frac{2(1+\beta^{-1})}{h(\tau)} \ln\left[\frac{2(1+\beta^{-1})}{h(\tau)} \ln n\right] + h(\tau)^{-1} \ln \alpha + C_2(\beta, \tau)$$

for some constant $C_2(\beta, \tau) > 0$, where $h(\tau)$ is defined as in (10).

Remark. The bounds in Propositions 4 and 5 have a similar form. In each case, the bound is of the form $n^{-(\beta+1)r}$ times a polynomial in $\ln n$, and the leading term in $s(\cdot)$ and $t(\cdot)$ are of the form $(1+\beta^{-1}) \ln n$ times a function of the threshold τ. The aspect ratio β of the target submatrix plays a critical role in both the size threshold and the probability bound. This reflects the dependence of the size of a $\lceil \beta k \rceil \times k$ submatrix on β. By contrast, the aspect ratio α of the sample matrix plays a secondary role, its logarithm appearing only in the constant term of $s(\cdot)$ and $t(\cdot)$.

5. Proof of Lemma 1 and Proposition 1

Proof of Lemma 1. Let $\tau > 0$ be fixed, and note that

$$\ln \phi_{n,\tau}(s) = \left(n + \frac{1}{2}\right) \ln n - \left(s + \frac{1}{2}\right) \ln s - \left(n - s + \frac{1}{2}\right) \ln(n-s) - \frac{\tau^2 s^2}{4} - \frac{1}{2} \ln 2\pi. \tag{11}$$

Differentiating $\ln \phi_{n,\tau}(s)$, with respect to s, yields

$$\frac{\partial \ln \phi_{n,\tau}(s)}{\partial s} = \frac{1}{2(n-s)} + \ln(n-s) - \frac{1}{2s} - \ln s - \frac{s\tau^2}{2}.$$

The last expression is negative when $2\tau^{-2} \ln n < s < 4\tau^{-2} \ln n$; we now consider the value of $\ln \phi_{n,\tau}(s)$ for s outside this interval. A straightforward calculation shows that for $0 < s \leq 2\tau^{-2} \ln n$,

$$\ln \phi_{n,\tau}(s) s \left(\ln(n - 2\tau^{-2} \ln n) - \frac{s\tau^2}{4} - \ln \ln n - \ln 2\tau^{-2}\right) - \frac{1}{2} \ln s - \frac{1}{2} \ln 2\pi,$$

which is positive when n is sufficiently large. In order to address the other extreme, note that, from (11), we have

$$\ln \phi_{n,\tau}(s) \leq s\left(\ln(n-s) - \frac{s\tau^2}{4} - \ln s\right) - \frac{1}{2} \ln s + (n + 1/2) \ln\left(\frac{n}{n-s}\right). \tag{12}$$

It is easy to check that the right-hand side of the above inequality is negative when $s > n - 2$. Considering separately the cases $s + 2 < n < (2\ln 2)^{-1} s \ln s$ and $n \geq (2\ln 2)^{-1} s \ln s$, one may upper bound the final term above by $(s \ln s)/2 + (\ln 2)/2$ and $2s + (\ln 2)/2$, respectively. Thus, for $s < n - 2$, we have

$$\ln \phi_{n,\tau}(s) \leq s\left(\ln(n-s) - \frac{s\tau^2}{4} - \ln s\right) - \frac{1}{2}\ln s + 2s + \frac{s \ln s}{2} + \frac{\ln 2}{2},$$

and, in particular, for $4\tau^{-2} \ln n \leq s < n - 2$,

$$\ln \phi_{n,\tau}(s) \leq s\left(2 - \frac{\ln s}{2}\right) - \frac{1}{2}\ln s + \frac{\ln 2}{2} < 0$$

when n (and therefore s) is sufficiently large. Thus for large n there exists a unique solution $s(n, \tau)$ of the equation $\phi_{n,\tau}(s) = 1$ with $s(n, \tau) \in (2\tau^{-2} \ln n, 4\tau^{-2} \ln n)$.

Taking logarithms of both sides of the equation $\phi_{n,\tau}(s) = 1$ and rearranging terms yields the expression

$$\frac{1}{2}\ln \frac{n}{n-s} + n \ln \frac{n}{n-s} - \left(s + \frac{1}{2}\right)\ln s + s\ln(n-s) - \frac{\tau^2 s^2}{4} = \frac{\ln 2\pi}{2}. \tag{13}$$

The argument above shows that the (unique) solution of this equation belongs to the interval $(2\tau^{-2} \ln n, 4\tau^{-2} \ln n)$, so we consider the case in which s and n/s tend to infinity with n. Dividing both sides of (13) by s yields

$$\ln(n-s) - \frac{s\tau^2}{4} - \ln s = -1 + \mathrm{O}\left(\frac{\ln s}{s}\right),$$

which, after adding and subtracting terms, can be rewritten in the equivalent form

$$\ln n - \frac{s\tau^2}{4} - \ln\ln n = \ln\left(\frac{s}{\ln n}\right) - \ln\left(\frac{n-s}{n}\right) - 1 + \mathrm{O}\left(\frac{\ln s}{s}\right). \tag{14}$$

For each $n \geq 1$, define $R(n)$ via the equation

$$s(n, \tau) = 4\tau^{-2} \ln n - 4\tau^{-2} \ln\ln n + R(n).$$

Plugging the last expression into (14), we find that $R(n) = \frac{4}{\tau^2}(1 - \ln \frac{4}{\tau^2}) + \mathrm{o}(1)$, and the result follows from the uniqueness of $s(n, \tau)$. □

Proof of Proposition 1. Fix $\tau > 0$. If $\lceil s(n, \tau)\rceil + r > n$, the bound (1) holds trivially; in the case of equality, it follows from a standard Gaussian tail bound when n is sufficiently large. Fix $n \geq 1$ for the moment, and suppose that $l = \lceil s(n, \tau)\rceil + r \leq n - 1$. By Markov's inequality and the definition of $\phi_{n,\tau}(\cdot)$,

$$\begin{aligned}P\big(K_\tau(W_n) \geq s(n,\tau)+r\big) &= P\big(K_\tau(W_n) \geq l\big)\\ &= P\big(\Gamma_l(n,\tau) \geq 1\big)\\ &\leq E\Gamma_l(n,\tau)\\ &\leq 2\phi_{n,\tau}^2(l) \leq 2\phi_{n,\tau}^2\big(s(n,\tau)+r\big).\end{aligned} \tag{15}$$

Let $\gamma = \mathrm{e}^{-\tau^2/4}$, and, to reduce notation, denote $s(n,\tau)$ by s_n. Under the constraint on r, a straightforward calculation shows that one can decompose the final term above as follows:

$$2\phi_{n,\tau}^2(s_n+r) = 2\phi_{n,\tau}^2(s_n)\gamma^{2rs_n}\big[A_n(r)B_n(r)C_n(r)D_n(r)\big]^2 \tag{16}$$

where

$$A_n(r) = \left(\frac{n-r-s_n}{n-s_n}\right)^{-n+r+s_n-1/2}, \qquad B_n(r) = \left(\frac{r+s_n}{s_n}\right)^{-s_n-1/2},$$
$$C_n(r) = \left(\frac{n-s_n}{r+s_n}\gamma^{s_n}\right)^r, \qquad D_n(r) = \gamma^{r^2}.$$

It is enough to bound the right-hand side of (16) as n increases, and $r = r(n)$ is such that $\lceil s(n,\tau)\rceil + r \leq n-1$. By definition, $\phi_{n,\tau}(s_n) = 1$, and

$$\max_{r\geq 1} \frac{2\gamma^{2rs_n}}{n^{-2r}(4\ln n/\tau^2)^{2r}} \to 0 \qquad \text{as } n\to\infty.$$

Thus it suffices to show that the product $A_n(r)B_n(r)C_n(r)D_n(r)$ is uniformly bounded in r. To begin, note that for any fixed $0<\delta<4$,

$$C_n(r)^{1/r} = \frac{n-s_n}{r+s_n}\gamma^{s_n} \leq \frac{n}{s_n}\gamma^{s_n} \leq \frac{4}{4-\delta}\mathrm{e}^{-1+\mathrm{o}(1)}.$$

The last term will be less than one when δ is sufficiently small and n is large. The term $B_n(r) \leq 1$ for each $r \geq 1$, so it only remains to show that $\max_{r\geq 1} A_n(r)\cdot D_n(r)$ is bounded as a function of n. A straightforward calculation shows that $\ln A_n(r) \leq r$, and consequently, $\ln(A_n(r)\cdot D_n(r)) \leq r - \frac{\tau^2 r^2}{4}$, a quadratic function of r that is bounded from above by $1/\tau^2$. □

Proof of Proposition 2. Let $k = \lceil 4\tau^{-2}\ln n\rceil + r \geq 3$. Following the argument in (15), we find that

$$\begin{aligned}P\big(K_\tau(W_n)\geq k\big) \leq E\Gamma_k(n,\tau) &= \binom{n}{k}^2 \exp\left\{-\frac{\tau^2k^2}{2}\right\}\\ &\leq \left(\frac{\mathrm{e}n}{k}\right)^{2k}\exp\left\{-\frac{\tau^2k^2}{2}\right\}\end{aligned}$$

$$
\begin{aligned}
&= \exp\left\{2k(1+\ln n-\ln k)-\frac{\tau^2k^2}{2}\right\}\\
&= \exp\left\{2k\left(1+\ln n-\ln k-\frac{\tau^2k}{4}\right)\right\}\\
&\le \exp\left\{2k\left(1-\ln k-\frac{\tau^2}{4}r\right)\right\}\\
&\le \exp\left\{-\frac{k\tau^2r}{2}\right\}\\
&\le n^{-2r}.
\end{aligned}
$$

The second inequality above makes use of the standard bound

$$\binom{n}{k} \le \left(\frac{\mathrm{e}n}{k}\right)^k.$$

The penultimate inequality follows from the fact that $k \ge 3$. □

6. Proof of Proposition 3

For any $k \times k$ submatrix U of the Gaussian random matrix W_n, it follows from standard arguments that $(k-1)^2G(U)$ has a χ^2 distribution with $(k-1)^2$ degrees of freedom. In order to bound the quantity $P(G(U) \le \tau)$, which arises in the analysis of $L_\tau(W_n)$, we require an initial result relating the right and left tails of the χ^2 distribution.

Lemma 2. *Suppose that $X \sim \chi^2_\ell$ for some $\ell \ge 3$. Then for $0 < t < \ell - 2$ we have*

$$P(X \le t) \le P(X \ge 2\ell - 4 - t).$$

Proof. Let f denote the density function of X and let $0 < t < \ell - 2$. Since

$$
\begin{aligned}
P(X \le t) &= \int_0^t f(s)\,\mathrm{d}s \quad \text{and}\\
P(X \ge 2\ell - 4 - t) &\ge \int_{2\ell-4-t}^{2\ell-4} f(s)\,\mathrm{d}s,
\end{aligned}
$$

it suffices to show that

$$\frac{f(s)}{f(2\ell-4-s)} \le 1 \qquad \text{for all } 0 < s < \ell - 2. \tag{17}$$

To this end, note that the ratio in (17) can be rewritten as follows:

$$\begin{aligned}\frac{f(s)}{f(2\ell-4-s)} &= \frac{s^{(\ell-2)/2}\mathrm{e}^{-s/2}}{(2\ell-4-s)^{(\ell-2)/2}\mathrm{e}^{-(2\ell-4-s)/2}} \\ &= \left[\left(1-\frac{2\ell-4-2s}{2\ell-4-s}\right)\mathrm{e}^{2(\ell-2-s)/(\ell-2)}\right]^{(\ell-2)/2} \\ &= \left[\left(1-\frac{1}{u}\right)\mathrm{e}^{2/(2u-1)}\right]^{(\ell-2)/2} \qquad \text{with } u=\frac{2\ell-4-s}{2\ell-4-2s}.\end{aligned} \tag{18}$$

As s tends to $\ell-2$, u tends to infinity, and therefore

$$\lim_{s\to(\ell-2)}\frac{f(s)}{f(2\ell-4-s)} = \lim_{u\to\infty}\left(1-\frac{1}{u}\right)\mathrm{e}^{2/(2u-1)} = 1.$$

Thus, it suffices to show that for $u\in(1,\infty)$, the final term in (18) is an increasing function of u. Differentiating with respect to u we find that

$$\frac{\mathrm{d}}{\mathrm{d}u}\left(1-\frac{1}{u}\right)\mathrm{e}^{2/(2u-1)} = \frac{(2u-1)^2-4(u-1)u}{u^2(2u-1)^2}\mathrm{e}^{2/(2u-1)} > 0$$

where the inequality follows from the fact that $u>1$. Inequality (17) follows immediately. □

Proof of Proposition 3. To begin, note that if X has a χ^2 distribution with ℓ degrees of freedom, then by a standard Chernoff bound,

$$P(X\geq r) \leq \min_{0<s<1/2}(1-2s)^{-\ell/2}\mathrm{e}^{-sr} = \left[\left(\frac{\ell}{r}\right)\mathrm{e}^{(r/\ell-1)}\right]^{-\ell/2}. \tag{19}$$

Let $0<\tau<1$ be fixed. Fix $n\geq 1$ for the moment and let $r\geq 1$ be such that $k=\lceil t(n,\tau)\rceil+r\leq n$, where $t(n,\tau)$ is defined as in the statement of Proposition 3. Let U be any $k\times k$ submatrix of W_n, and let $\ell=(k-1)^2$. As noted above, the random variable $\ell G(U)$ has a χ^2 distribution with ℓ degrees of freedom, so by Lemma 2 and inequality (19),

$$\begin{aligned}P\big(G(U)\leq\tau\big) &= P\big(\ell G(U)\leq\ell\tau\big)\leq P\big(\ell G(U)\geq(2-\tau)\ell-4\big) \\ &\leq \exp\left\{-\frac{\ell}{2}\left[\frac{(2-\tau)\ell-4}{\ell}-1+\ln\frac{\ell}{(2-\tau)\ell-4}\right]\right\} \\ &= \exp\left\{-\frac{\ell}{2}[(1-\tau)-\ln(2-\tau)]\right\}\exp\left\{\left[2+\frac{\ell}{2}\ln\left(1-\frac{4}{\ell(2-\tau)}\right)\right]\right\} \\ &\leq \exp\left\{-\frac{\ell}{2}[(1-\tau)-\ln(2-\tau)]\right\}\exp\left\{2-\frac{2}{2-\tau}\right\}.\end{aligned} \tag{20}$$

The second term in the last display is, at most, e. It follows from a first moment argument that

$$\begin{aligned} P\big(L_\tau(W_n) \geq k\big) &\leq \binom{n}{k}^2 P\big(G(U) \leq \tau\big) \\ &\leq \mathrm{e} \binom{n}{k}^2 q^{(k-1)^2} \\ &\leq \mathrm{e} \binom{n}{k-1}^2 q^{(k-1)^2} \cdot n^2, \end{aligned}$$

where

$$q = \exp\big\{\tfrac{1}{2}[-(1-\tau) + \ln(2-\tau)]\big\}.$$

The quantity $h(\tau) = (1-\tau) - \ln(2-\tau) \geq 0$ as $0 < \tau < 1$. Define

$$\tau_0 = \sqrt{h(\tau)} = \sqrt{(1-\tau) - \ln(2-\tau)},$$

and note that

$$\begin{aligned} k &= \left\lceil \frac{4}{h(\tau)} \ln n - \frac{4}{h(\tau)} \ln\left(\frac{4}{h(\tau)} \ln n\right) + \frac{4}{h(\tau)} \right\rceil + 2 + r \\ &\geq s(n, \tau_0) + 2 + r, \end{aligned}$$

where $s(n, \tau_0)$ is defined as in Lemma 1. Following the argument after inequality (15) in the proof of Proposition 1, and using the monotonicity of ϕ_{n,τ_0} for sufficiently large n, we find that

$$\begin{aligned} \binom{n}{k-1}^2 q^{(k-1)^2} &= \binom{n}{k-1}^2 \mathrm{e}^{-\tau_0^2 (k-1)^2/2} \\ &\leq 2\phi^2_{n,\tau_0}(k-1) \\ &\leq 2\phi^2_{n,\tau_0}\big(s(n,\tau_0) + r + 1\big) \\ &\leq 2\mathrm{e}^{2/\tau_0^2} \left(\frac{\ln n}{\tau_0^2}\right)^{2r+2} n^{-2r-2} \\ &= 2\mathrm{e}^{2/h(\tau)} \left(\frac{\ln n}{h(\tau)}\right)^{2r+2} n^{-2r-2}. \end{aligned}$$

The result then follows from (21). □

7. Proof of Theorem 1

In what follows we make use of standard bounds on the tails of the Gaussian distribution, namely that $(3s)^{-1}\mathrm{e}^{-s^2/2} \leq 1 - \Phi(s) \leq s^{-1}\mathrm{e}^{-s^2/2}$ for $s \geq 3$. The proof of Theorem 1 is based on several preliminary results. The first result bounds the ratio of the variance of $\Gamma_k(\tau, n)$ and the square of its expected value, a quantity that later arises from an application of Chebyshev's inequality.

Lemma 3. *Fix $\tau > 0$. There exist integers $n_0, k_0 \geq 1$ and a positive constant C depending on τ but independent of k and n, such that for any $n \geq n_0$ and any $k \geq k_0$,*

$$\frac{\operatorname{Var}\Gamma_k(\tau, n)}{(E\Gamma_k(\tau, n))^2} \leq Ck^4 \sum_{l=1}^{k} \sum_{r=1}^{k} \frac{\binom{k}{l}\binom{n-k}{k-l}}{\binom{n}{k}} \frac{\binom{k}{r}\binom{n-k}{k-r}}{\binom{n}{k}} \exp\left\{\frac{rl\tau^2}{2}\left(1 + \frac{k^2 - rl}{k^2 + rl}\right)\right\}. \tag{21}$$

Proof. Let $\mathcal{S}_k$ denote the collection of all $k \times k$ submatrices of W_n. It is clear that

$$E\Gamma_k(n, \tau) = \sum_{U \in \mathcal{S}_k} P\bigl(F(U) > \tau\bigr) = \binom{n}{k}^2 \bigl(1 - \Phi(k\tau)\bigr). \tag{22}$$

In a similar fashion, we have

$$E\Gamma_k^2(n, \tau) = \sum_{U_i, U_j \in \mathcal{S}_k} P\bigl(F(U_i) > \tau \text{ and } F(U_j) > \tau\bigr).$$

Note that the joint probability in the last display depends only on the overlap between the submatrices U_i and U_j. For $1 \leq r, l \leq k$ define

$$G(r, l) = P\bigl(F(U) > \tau \text{ and } F(V) > \tau\bigr),$$

where U and V are two fixed $k \times k$ submatrices of W having r rows and l columns in common. Note that $G(r, l) = 0$ if $2k - r > n$ or $2k - l > n$. A straightforward counting argument shows that

$$E\Gamma_k^2(n, \tau) = \sum_{r=0}^{k} \sum_{l=0}^{k} \binom{n}{k}^2 \binom{k}{r} \binom{n-k}{k-r} \binom{k}{l} \binom{n-k}{k-l} G(r, l).$$

In particular,

$$\begin{aligned}\frac{\operatorname{Var}\Gamma_k(n, \tau)}{(E\Gamma_k(n, \tau))^2} &= \sum_{r=0}^{k} \sum_{l=0}^{k} \frac{\binom{k}{l}\binom{n-k}{k-l}}{\binom{n}{k}} \frac{\binom{k}{r}\binom{n-k}{k-r}}{\binom{n}{k}} \cdot \frac{G(r, l)}{(1 - \Phi(k\tau))^2} - 1 \\ &\leq \sum_{r=1}^{k} \sum_{l=1}^{k} \frac{\binom{k}{l}\binom{n-k}{k-l}}{\binom{n}{k}} \frac{\binom{k}{r}\binom{n-k}{k-r}}{\binom{n}{k}} \left[\frac{G(r, l)}{(1 - \Phi(k\tau))^2} - 1\right].\end{aligned}$$

Here we have used the fact that $\binom{k}{l}\binom{n-k}{k-l}/\binom{n}{k}$ is a probability mass function, and that $G(r, 0)$ and $G(0, l)$ are either equal to zero or equal to $(1 - \Phi(k\tau))^2$. When $k\tau \geq 3$ we have $(1 - \phi(k\tau))^2 \geq (3k\tau)^{-2} \mathrm{e}^{-k^2\tau^2}$. It therefore suffices to show that for $1 \leq r, l \leq k$, such that $2k - r \leq n$

and $2k - l \leq n$, one has

$$G(r,l) \leq Ck^2 \exp\left\{-k^2\tau^2 + \frac{rl\tau^2}{2}\left(1 + \frac{k^2 - rl}{k^2 + rl}\right)\right\}, \tag{23}$$

where $C > 0$ depends on τ but is independent of k and n. Inequality (23) is readily established when $r = l = k$, so we turn our attention to bounding the quantity $G(r,l)$ when it is positive and $1 \leq rl < k^2$. In this case

$$G(r,l) = \frac{\sqrt{rl}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-rlt^2/2} P\left(F(U \cap V^c) \geq \frac{k^2\tau - rlt}{\sqrt{k^2 - rl}}\right)^2 \mathrm{d}t,$$

where U, V are submatrices of W_n having r rows and l columns in common. Let $\overline{\Phi}(x) = 1 - \Phi(x)$. Note that $G(r,l) = D_0 + D_1$ where

$$D_0 = \frac{\sqrt{rl}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-rlt^2/2} \overline{\Phi}^2\left(\frac{k^2\tau - rlt}{\sqrt{k^2 - rl}}\right) I\{k^2\tau - rlt < 1\}\, \mathrm{d}t \tag{24}$$

and

$$D_1 = \frac{\sqrt{rl}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-rlt^2/2} \overline{\Phi}^2\left(\frac{k^2\tau - rlt}{\sqrt{k^2 - rl}}\right) I\{k^2\tau - rlt \geq 1\}\, \mathrm{d}t. \tag{25}$$

Consider first the term D_1 defined in (25). As $rl \neq k^2$ and $k^2\tau - rlt \geq 1$, the normal tail bound yields

$$\begin{aligned}\overline{\Phi}\left(\frac{k^2\tau - rlt}{\sqrt{k^2 - rl}}\right) &\leq \frac{\sqrt{k^2 - rl}}{\sqrt{2\pi}(k^2\tau - rlt)} \exp\left\{-\frac{(k^2\tau - rlt)^2}{2(k^2 - rl)}\right\} \\ &= \mathrm{O}\left(\sqrt{k^2 - rl}\right) \exp\left\{-\frac{(k^2\tau - rlt)^2}{2(k^2 - rl)}\right\}.\end{aligned}$$

Plugging the last expression into (25), the exponential part of the resulting integrand is

$$-\frac{(k^2\tau - rlt)^2}{(k^2 - rl)} - \frac{rlt^2}{2},$$

which (after lengthy but straightforward algebra) can be expressed as

$$-k^2\tau + \frac{rl\tau^2}{2}\left(1 + \frac{k^2 - rl}{k^2 + rl}\right) - \frac{rl(k^2 + rl)}{2(k^2 - rl)}\left((\tau - t) + \tau\left(\frac{k^2 - rl}{k^2 + rl}\right)\right)^2.$$

It then follows that

$$\begin{aligned}D_1 &\leq \mathrm{O}(k^2 - rl) \exp\left\{-k^2\tau^2 + \frac{rl\tau^2}{2}\left(1 + \frac{k^2 - rl}{k^2 + rl}\right)\right\} \\ &\quad \times \sqrt{\frac{k^2 - rl}{k^2 + rl}} \times \int_{-\infty}^{\infty} \sqrt{\frac{rl(k^2 + rl)}{k^2 - rl}} \exp\left\{-\frac{rl(k^2 + rl)}{2(k^2 - rl)}\left(\tau - t + \frac{\tau(k^2 - rl)}{k^2 + rl}\right)^2\right\} \mathrm{d}t.\end{aligned}$$

The term preceding the integral is less than one, and the integral is equal to one. Thus D_1 is less than the right-hand side of (23).

We next consider the term D_0 defined in (24). Note that $k^2\tau - rlt < 1$ is equivalent to $t > (k^2\tau - 1)/rl$, and therefore

$$D_0 \le \int_{(k^2\tau-1)/rl}^{\infty} \frac{\sqrt{rl}}{\sqrt{2\pi}} \mathrm{e}^{-rlt^2/2}\,\mathrm{d}t = \overline{\Phi}\biggl(\frac{k^2\tau-1}{\sqrt{rl}}\biggr) \le \frac{k\sqrt{rl}}{\sqrt{2\pi}(k^2\tau-1)} \mathrm{e}^{-(k^2\tau-1)^2/(2rl)-\ln k}.$$

Comparing the last term above with (23), it suffices to show that when k is sufficiently large,

$$\frac{(k^2\tau-1)^2}{2rl} + \ln k \ge \biggl(k^2 - \frac{rl}{2}\biggr)\tau^2$$

or, equivalently,

$$(k^2 - rl)^2\tau^2 - 2k^2\tau + 1 + 2rl\ln k \ge 0. \tag{26}$$

Suppose first that $rl \ge k^2 - k/\sqrt{\ln k}$. In this case, the left-hand side of the expression above is, at least,

$$-2k^2\tau + 1 + 2rl\ln k \ge -2k^2\tau + 1 + 2\bigl(k^2 - k/\sqrt{\ln k}\bigr)\ln k > 0$$

when k is sufficiently large. Suppose now that $k^2 - rl > k/\sqrt{\ln k}$. As a quadratic function of τ, the left-hand side of (26) takes its minimum at $\tau = k^2/(k^2 - rl)^2$, and the corresponding value is $rl[-2k^2 + rl + 2(k^2 - rl)^2\ln k]/(k^2 - rl)^2$. In this case, the assumption $k^2 - rl > k/\sqrt{\ln k}$ implies

$$-2k^2 + rl + 2(k^2 - rl)^2\ln k > rl > 0.$$

This establishes (26) and complete the proof. □

Lemma 4. *Let $\tau > 0$ be fixed. When k is sufficiently large, for every integer n satisfying the condition*

$$k \le \frac{4}{\tau^2}\ln n - \frac{4}{\tau^2}\ln\biggl(\frac{4}{\tau^2}\ln n\biggr) - \frac{12\ln 2}{\tau^2} \tag{27}$$

we have the bound

$$\frac{\operatorname{Var}\Gamma_k(\tau, n)}{(E\Gamma_k(\tau, n))^2} \le k^{-2}.$$

Remark. For the proof of Theorem 1, it is enough to show that the sum over k of the ratio above is finite, and, for this purpose, the upper bound k^{-2} is sufficient.

Proof of Lemma 4. Let n satisfy condition (27). By Lemma 3, it suffices to show that

$$k^4 \sum_{l=1}^{k}\sum_{r=1}^{k} \frac{\binom{k}{l}\binom{n-k}{k-l}}{\binom{n}{k}} \frac{\binom{k}{r}\binom{n-k}{k-r}}{\binom{n}{k}} \exp\biggl\{\frac{rl\tau^2}{2}\biggl(1 + \frac{k^2 - rl}{k^2 + rl}\biggr)\biggr\} \le k^{-2}. \tag{28}$$

In order to establish (28), we will show that each term in the sum is less than k^{-8}. To begin, note that

$$\frac{\binom{k}{l}\binom{n-k}{k-l}}{\binom{n}{k}} \leq \frac{\binom{k}{l}k^l(n-k)^{k-l}}{(n-k)^k} = \binom{k}{l} k^l (n-k)^{-l},$$

and that $(n-k)^{-l} = \mathrm{O}(n^{-l})$ when $l \leq k = \mathrm{O}(n^{1/2})$. Thus for some constant $C > 0$,

$$\frac{\binom{k}{l}\binom{n-k}{k-l}}{\binom{n}{k}} \frac{\binom{k}{r}\binom{n-k}{k-r}}{\binom{n}{k}} \leq C \binom{k}{r}\binom{k}{l} k^{r+l} n^{-(r+l)}.$$

Rewriting (27) as $\ln n \geq \frac{\tau^2 k}{4} + \ln(\frac{4}{\tau^2}\ln n) + 3\ln 2$ yields the bound

$$\begin{aligned} &n^{-(r+l)} \exp\left\{\frac{rl\tau^2}{2}\left(1 + \frac{k^2 - rl}{k^2 + rl}\right)\right\} \\ &\quad \leq \mathrm{e}^{-3(r+l)\ln 2}\left(\frac{4}{\tau^2}\ln n\right)^{-(r+l)} \exp\left\{\frac{\tau^2}{2}\left(rl\frac{2k^2}{k^2+rl} - \frac{k}{2}(r+l)\right)\right\}. \end{aligned}$$

Combining the last three displays, and using the fact that $k \leq \frac{4}{\tau^2}\ln n$ by assumption, it suffices to show that

$$\binom{k}{r}\binom{k}{l} \mathrm{e}^{-3(r+l)\ln 2} \exp\left\{\frac{\tau^2}{2}\left(rl\frac{2k^2}{k^2+rl} - \frac{k}{2}(r+l)\right)\right\} \leq k^{-8}. \tag{29}$$

In order to establish (29), we consider two cases for $r+l$. Suppose first that $r + l \leq \frac{3k}{4}$. By elementary arguments,

$$\binom{k}{r}\binom{k}{l} \leq \binom{2k}{r+l} \leq (2k)^{r+l} \quad \text{and} \quad rl\frac{2k^2}{k^2+rl} \leq \frac{(r+l)^2}{4}\frac{2k^2}{k^2+rl} \leq \frac{(r+l)^2}{2}.$$

It follows from these inequalities that

$$\begin{aligned} &\binom{k}{r}\binom{k}{l} \exp\left\{\frac{\tau^2}{2}\left[rl\frac{2k^2}{k^2+rl} - \frac{k}{2}(r+l)\right]\right\} \\ &\quad \leq \exp\left\{\frac{\tau^2}{2}\left[\frac{(r+l)^2}{2} - \frac{k}{2}(r+l)\right] + (r+l)\ln 2k\right\} \\ &\quad = \exp\left\{\frac{\tau^2(r+l)}{2}\left[\frac{(r+l)}{2} - \frac{k}{2} + \frac{2\ln 2k}{\tau^2}\right]\right\} \\ &\quad \leq \exp\left\{\frac{\tau^2(r+l)}{2}\left[\frac{3k}{8} - \frac{k}{2} + \frac{2\ln 2k}{\tau^2}\right]\right\}. \end{aligned}$$

As the exponent above is negative when k is sufficiently large, (29) follows. Suppose now that $r + l \geq \frac{3k}{4}$. From the simple bounds $r + l \geq 2\sqrt{rl}$ and $k^2 + rl \geq 2\sqrt{k^2 rl}$, we find that

$$rl\frac{2k^2}{k^2 + rl} - \frac{k}{2}(r + l) \leq \frac{2rlk^2}{2\sqrt{k^2 rl}} - k\sqrt{rl} = 0,$$

and it suffices to bound the initial terms in (29). But, clearly,

$$\binom{k}{r}\binom{k}{l}\mathrm{e}^{-3(r+l)\ln 2} \leq 2^{2k} \cdot 2^{-9k/4},$$

which is less than k^{-8} when k is sufficiently large. □

Proof of Theorem 1. Proposition 1 and the Borel–Cantelli lemma imply that eventually, almost surely, $K_\tau(W_n) \leq \lceil s(n, \tau)\rceil + 1$. Thus, we only need to establish an almost sure lower bound on $K_\tau(W_n)$. To this end, define functions

$$f(n) = \frac{4}{\tau^2}\ln n - \frac{4}{\tau^2}\ln\left(\frac{4}{\tau^2}\ln n\right) - \frac{12\ln 2}{\tau^2} \quad \text{and} \quad g(k) = \min\{n \geq 1, \lfloor f(n)\rfloor = k\}$$

for integers $n \geq 1$ and $k \geq 1$, respectively. It is easy to see that $f(n)$ is strictly increasing for large values of n, and clearly $f(n)$ tends to infinity as n tends to infinity. A straightforward argument shows that $g(k)$ has the same properties. Thus for every sufficiently large integer n, there exists a unique integer $k = k(n)$ such that $g(k) \leq n < g(k + 1)$.

Fix $m \geq 1$ and consider the event A_m that for some $n \geq m$ the random variable $K_\tau(W_n)$ is less than the lower bound specified in the statement of the theorem. More precisely, define

$$A_m = \bigcup_{n\geq m}\left\{K_\tau(W_n) \leq s(n, \tau) - \frac{12\ln 2}{\tau^2} - \frac{4}{\tau^2} - 3\right\}.$$

To establish the lower bound, it suffices to show that $P(A_m) \to 0$ as $m \to \infty$. To begin, note that when m is large,

$$A_m \subseteq \bigcup_{k\geq\lfloor f(m)\rfloor}\ \bigcup_{g(k)\leq n<g(k+1)}\left\{K_\tau(W_n) \leq s(n, \tau) - \frac{12\ln 2}{\tau^2} - \frac{4}{\tau^2} - 4\right\}.$$

Fix $n \geq m$ sufficiently large, and let $k = k(n)$ be the unique integer such that $g(k) \leq n < g(k+1)$. The definition of $g(k)$ and the monotonicity of $f(\cdot)$ ensures that $k = \lfloor f(g(k))\rfloor \leq f(n) < k + 1$. In conjunction with the definition of $f(n)$ and Lemma 1, this inequality implies that

$$\begin{aligned}1 = k + 1 - k > f(n) - \lfloor f(g(k))\rfloor &\geq f(n) - f(g(k))\\ &= s(n, \tau) - s(g(k), \tau) + \mathrm{o}(1),\end{aligned}$$

and therefore $s(n, \tau) < s(g(k), \tau) + 1 + \mathrm{o}(1)$. Define

$$r(k) = \left\lfloor s(g(k), \tau) - \frac{12\ln 2}{\tau^2} - \frac{4}{\tau^2}\right\rfloor.$$

From the bound on $s(n,\tau)$ above and the fact that $K_\tau(W_{g(k)}) \le K_\tau(W_n)$, we have

$$\left\{K_\tau(W_n) \le s(n,\tau) - \frac{12\ln 2}{\tau^2} - \frac{4}{\tau^2} - 3\right\} \subseteq \left\{K_\tau\left(W_{g(k)}\right) \le r(k) - 1 + \mathrm{o}(1)\right\}$$
$$\subseteq \left\{K_\tau\left(W_{g(k)}\right) \le r(k) - 1\right\},$$

where the last relation makes use of the fact that K_τ and $r(k)$ are integers. Thus we find that

$$A_m \subseteq \bigcup_{k \ge \lfloor f(m) \rfloor} \left\{K_\tau\left(W_{g(k)}\right) \le r(k) - 1\right\}.$$

Consider the events above. For fixed k,

$$P\left(K_\tau\left(W_{g(k)}\right) \le r(k) - 1\right) = P\left(\Gamma_{r(k)}(\tau, g(k)) = 0\right) \le \frac{\operatorname{Var}\Gamma_{r(k)}(\tau, g(k))}{(E\Gamma_{r(k)}(\tau, g(k)))^2} \tag{30}$$

where we have used the fact that for a non-negative integer-valued random variable X

$$P(X = 0) \le P(|X - EX| \ge EX) \le \frac{\operatorname{Var} X}{(EX)^2},$$

by Chebyshev's inequality. As $r(k) \le f(g(k))$, Lemma 4 ensures that the final term in (30) is less than k^{-2}, and the Borel–Cantelli lemma then implies that $P(A_m) \to 0$ as $m \to \infty$. This completes the proof of Theorem 1. □

Acknowledgements

The authors would like to thank Andrey Shabalin for his help in clarifying the connections between the work described here and results in random matrix theory. The work presented in this paper was supported in part by NSF Grants DMS 04-06361 and DMS 09-07177.

References

[1] Alon, N. and Naor, A. (2006). Approximating the cut-norm via Grothendieck's inequality. *SIAM J. Comput.* **35** 787–803 (electronic). MR2203567

[2] Anderson, G.W., Guionnet, A. and Zeitouni, O. (2010). *An Introduction to Random Matrices. Cambridge Studies in Advanced Mathematics* **118**. Cambridge: Cambridge Univ. Press. MR2760897

[3] Bollobás, B. and Erdős, P. (1976). Cliques in random graphs. *Math. Proc. Cambridge Philos. Soc.* **80** 419–427. MR0498256

[4] Dawande, M., Keskinocak, P., Swaminathan, J.M. and Tayur, S. (2001). On bipartite and multipartite clique problems. *J. Algorithms* **41** 388–403. MR1869258

[5] Geman, S. (1980). A limit theorem for the norm of random matrices. *Ann. Probab.* **8** 252–261. MR0566592

[6] Koyuturk, M., Szpankowski, W. and Grama, A. (2004). Biclustering gene-feature matrices for statistically significant dense patterns. In *Proceedings of the* 2004 *IEEE Computational Systems Bioinformatics Conference* 480–484. IEEE Computer Society, Technical Committee on Bioinformatics.

[7] Matula, D. (1976). The largest clique size in a random graph. Technical Report CS 7608, Southern Methodist Univ.

[8] Paul, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica* **17** 1617–1642. MR2399865

[9] Scheffé, H. (1999). *The Analysis of Variance*. *Wiley Classics Library*. New York: Wiley. MR1673563

[10] Sun, X. and Nobel, A. (2006). Significance and recovery of block structures in binary matrices with noise. In *Proceedings of the* 19*th Conference on Learning Theory*. *Lecture Notes in Computer Science* **4005** 109–122. Berlin: Springer. MR2277922

[11] Sun, X. and Nobel, A.B. (2008). On the size and recovery of submatrices of ones in a random binary matrix. *J. Mach. Learn. Res.* **9** 2431–2453. MR2460888

Received June 2010 and revised August 2011