On the concentration of eigenvalues of random symmetric matrices

Michael Krivelevich* Van H. Vu †

Abstract

We prove that the few largest (and most important) eigenvalues of random symmetric matrices of various kinds are very strongly concentrated. This strong concentration enables us to compute the means of these eigenvalues with high precision. Our approach uses Talagrand's inequality and is very different from standard approaches.

1 Introduction

In this paper we consider the eigenvalues of random symmetric matrices whose diagonal and upper diagonal entries are independent random variables. Our goal is to study few largest/smallest eigenvalues of such a matrix. Let us begin with a version of Wigner's famous semi-circle law [12], due to Arnold [1, 6], which describes the limiting behavior of the bulk of the spectrum of a random matrix of this type.

Semi-circle law. *For $1 \le i \le j \le n$ let a_{ij} be real value random variables such that all a_{ij}, $i < j$ have the same distribution and all a_{ii} have the same distribution. Assume that all central moments of the a_{ij} are finite and put $\sigma^2 = \sigma^2(a_{ij})$. For $i < j$ set $a_{ji} = a_{ij}$ and let A_n denote the random matrix $(a_{ij})_1^n$. Finally, denote by $W_n(x)$ the number of eigenvalues of A_n not larger than x, divided by n. Then*

$$\lim_{n \to \infty} W_n(x 2\sigma\sqrt{n}) = W(x) \ ,$$

in distribution, where $W(x) = 0$ if $x \le -1$, $W(x) = 1$ if $x \ge 1$ and $W(x) = \frac{2}{\pi}\int_{-1}^{x}(1-x^2)^{1/2}dx$ if $-1 \le x \le 1$.

The semi-circle law gives only a limit distribution and does not tell anything about the behavior of the largest/smallest (and usually most important) eigenvalues. These eigenvalues were studied in several papers [4, 3, 8, 9]. The method used in these papers is to estimate the expectation of the trace of a high power of the matrix. This frequently leads to a sharp upper bound on the largest eigenvalue (see Section 2).

Given a symmetric matrix A, we denote by $\delta_1(A) \ge \delta_2(A) \ge \ldots \ge \delta_n(A)$ the eigenvalues of A. Furthermore, let $\lambda_1(A) = \max_{i=1}^n(|\delta_i A)|) = \max(|\delta_1(A)|, |\delta_n(A)|)$ and $\lambda_2(A) = \max(|\delta_2(A)|, |\delta_n(A)|)$.

*Department of Mathematics, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel. Email: krivelev@math.tau.ac.il.

†Microsoft Research, 1 Microsoft Way, Redmond, WA 98052, USA. E-mail: vanhavu@microsoft.com.

The purpose of this paper is to prove large deviation bounds for $\lambda_1, \lambda_2, \delta_1, \delta_2$ and δ_n. We believe that these results are of interest for a number of reasons. The first is that our results are obtained under a very general assumption on the distribution of the entries of a random symmetric matrix. Secondly, our large deviation bounds turn out to be very strong. Moreover, they are sharp, up to a constant in the exponent, in a certain deviation range. Also, our method appears to be new; it makes a novel application of the recent and powerful inequality of Talagrand [10]. Finally, since bounds on the largest eigenvalues of a symmetric random matrix are widely used in many applications in Combinatorics and Theoretical Computer Science, we believe that our results have a potential in these areas. As an example of such an application, we would like to mention a paper [5] of the present authors, where a version of our theorems has been used to design approximation algorithms with expected polynomial running time for such important computational problems as finding the chromatic number and the independence number of a graph.

Our first result involves the following general model. Let a_{ij} $(1 \leq i \leq j \leq n)$ be independent random variables, with absolute value at most 1. A symmetric random matrix A is obtained by defining $a_{ji} = a_{ij}$ for all $i < j$.

Theorem 1 *There are positive constants c and K such that for any $t > K$,*

$$Pr[|\lambda_1(A) - E(\lambda_1(A))| \geq t] \leq e^{-ct^2} \ .$$

The same result holds for both $\delta_1(A)$ and $\delta_n(A)$.

The bound in Theorem 1 is sharp, up to the constant c, when t is sufficiently large. The surprising fact about this theorem is that it requires basically no knowledge about the distributions of the a_{ij}.

Our second theorem provides a large deviation result for the second largest eigenvalue of a symmetric random matrix A, under the additional assumption that all non-diagonal entries of A have the same expectation $p > 0$.

Theorem 2 *For every constant $p > 0$ there exists constants $c_p, K_p > 0$ so that the following holds. If in addition to the conditions of Theorem 1, the random variables a_{ij}, $1 \leq i < j \leq n$ satisfy $E[a_{ij}] = p$, then for all $t > K_p$,*

$$Pr[\lambda_2(A) - E(\lambda_2(A))| \geq t] \leq e^{-c_p t^2} \ .$$

The same result holds for $\delta_2(A)$.

One particular application of the above theorem arises when all diagonal entries of A are 0, and each non-diagonal entry of A is a Bernoulli random variable with parameter p, i.e. $Pr[a_{ij} = 1] = p$, $Pr[a_{ij} = 0] = 1 - p$. In this case A can be viewed as the adjacency matrix of the *random graph* $G(n, p)$. Thus Theorem 2 provides in this case a large deviation result for the second eigenvalue of a random graph. In fact, for this special case, Theorem 2 can be extended for p decreasing in n (see Section 5).

The rest of the paper is organized as follows. In the next section, we collect some information about the expectations of the eigenvalues in concern. More interesting, it turns out that our theorems can sometimes be used to estimate these expectations. The proofs of Theorems 1 and 2 appear in Sections 3 and 4, respectively. We end with Section 5, which contains few remarks and open questions.

In what follows, a matrix is always symmetric and of order n, if not otherwise specified. We assume that n tends to infinity and the asymptotic notations (such as o, O, etc) are understood under this assumption. The letter c denotes a positive constant, whose value may vary in different occurrences. Bold lower case letters such as $\mathbf{x}, \mathbf{y}$ denote vectors in R^n and $\mathbf{xy}$ is the inner product of $\mathbf{x}$ and $\mathbf{y}$. Given a matrix A, $\mathbf{x}A\mathbf{y}$ is the inner product of $\mathbf{x}$ and $A\mathbf{y}$. $\mathbf{1}$ is the all one vector.

2 Expectations

In this section, we present several results about the expectation of the relevant eigenvalues. We also show that our theorems can be used to determine these expectations in some cases.

Let a_{ij}, $i \le j$ be independent random variables bounded in their absolute values by 1. Assume that for $i < j$, the a_{ij} have common expectation p and variance σ^2. Furthermore, assume that $E[a_{ii}] = \nu$ for all i. Füredi and Komlós ([3], Theorem 1), showed that if $p > 0$ then

$$E[\lambda_1(A)] = (n-1)p + \nu + \sigma^2/p + o(1) \ . \tag{1}$$

Also, in this case under a weaker assumption $VAR[a_{ij}] \le \sigma^2$ for all $1 \le i \le j \le n$ the argument of Füredi and Komlós gives:

$$E[\lambda_2(A)] \le 2\sigma\sqrt{n} + O(n^{1/3}\log n) \ . \tag{2}$$

The situation changes when $p = 0$. In the same paper, Füredi and Komlós (implicitly) showed that in this case (again assuming only $VAR[a_{ij}] \le \sigma^2$)

$$E[\lambda_1(A)] \le 2\sigma\sqrt{n} + O(n^{1/3}\log n) \ . \tag{3}$$

Füredi and Komlós also claimed that if $Var[a_{ij}] = \sigma^2$ then with probability tending to 1, $\lambda_1(A) \ge 2\sigma\sqrt{n} + O(n^{1/3}\log n)$.

Using our Theorem 1, we first show that a statement slightly weaker than (1) holds under a more general assumption that the variances are upper bounded by σ^2, but are not necessarily equal. Next, we prove a lower bound stronger than that stated by Füredi and Komlós.

Corollary 2.1 *If all entries a_{ij} of the random symmetric matrix $A = (a_{ij})$ are bounded in absolute value by 1, and all non-diagonal entries have common expectation $p > 0$, then*

$$E[\lambda_1(A)] = np + O(\sqrt{n}) \ .$$

Proof. For each entry a_{ij}, one can define a random variable c_{ij}, satisfying $|c_{ij}| \le 1$, $E[c_{ij}] = 0$, $VAR[c_{ij}] = 1 - VAR[a_{ij}]$. Let now $b_{ij} = a_{ij} - c_{ij}$. Then clearly $E[b_{ij}] = p$, $VAR[b_{ij}] = 1$. Denote $B = (b_{ij})$, $C = c_{ij}$, then $A = B + C$. Hence $\lambda_1(A) \le \lambda_1(B) + \lambda_1(C)$. Applying (1), (3) and Theorem 1, we obtain:

$$\begin{aligned} Pr[\lambda_1(B) \le np + O(1)] &\ \ge\ \frac{3}{4}, \\ Pr[\lambda_1(C) \le O(\sqrt{n})] &\ \ge\ \frac{3}{4}, \end{aligned}$$

and thus $Pr[\lambda_1(A) = np + O(\sqrt{n})] \geq 1/2$. Invoking Theorem 1 once again, we get the desired result. $\square$

By the same argument, one can show that if $\sigma = \omega(n^{-1/2})$, then $E(\delta_n) = 2\sigma n^{-1/2}(1 + o(1))$.

Corollary 2.2 *If all entries a_{ij} of the random symmetric matrix $A = (a_{ij})$ have common expectation 0 and variance σ^2, then*

$$E[\lambda_1(A)] \geq 2\sigma n^{1/2} + O(\log^{1/2} n) \ .$$

Consequently, with probability tending to 1,

$$\lambda_1(A) \geq 2\sigma n^{1/2} + O(\log^{1/2} n) \ .$$

Proof. For the sake of simplicity, we assume $\sigma = 1/2$. Furthermore, set $\mu = n^{1/2}$, $k = \lceil \mu \log^{1/2} n \rceil$ and $x = a \log^{1/2} n$, where a is a positive constant chosen so that the following two inequalities hold:

$$\mu^k / k^{5/2} \geq 2(\mu - x/2)^k \tag{4}$$

$$\sum_{t=\frac{a}{2}\log^{1/2} n}^{\infty} e^{2t \log^{1/2} n - ct^2} = o(1), \tag{5}$$

where c is the constant in Theorem 1. Without loss of generality, we assume that k is an even integer and let X be the trace of A^k. It is trivial that $E[X] \leq nE[\lambda_1^k]$. On the other hand, a simple counting argument (see [3]) shows that

$$E[X] \geq \frac{1}{(k/2)+1} \binom{k}{k/2} \sigma^k n(n-1)\dots(n-(k/2)).$$

It follows that

$$E[\lambda_1^k] \geq \frac{1}{(k/2)+1} \binom{k}{k/2} \sigma^k (n-1)\dots(n-(k/2)) \geq \mu^k / k^{5/2} \, . \tag{6}$$

Assume, for contradiction, that $E(\lambda_1) \leq \mu - x$. It follows from this assumption that

$$E[\lambda_1^k] \leq (\mu - x/2)^k + \sum_{t=x/2}^{\infty} (\mu - x + (t+1))^k Pr[\lambda_1 \geq \mu - x + t] \, . \tag{7}$$

By Theorem 1, $Pr(\lambda_1 \geq \mu - x + t) \leq e^{-ct^2}$ for all $t \geq x/2$. Thus (4),(6) and (7) imply

$$\sum_{t=x/2}^{\infty} (\mu - x + (t+1))^k e^{-ct^2} \geq \mu^k / k^{5/2} - (\mu - x/2)^k \geq (\mu - x/2)^k. \tag{8}$$

Since $(\mu - x + (t+1))^k / (\mu - x/2)^k \leq e^{(1+o(1))tk/\mu} = e^{(1+o(1))t \log^{1/2} n}$, (5) and (8) imply a contradiction, and this completes the proof. $\square$

To end this section, let us mention few recent results of Sinai and Soshnikov. In [8], Sinai and Soshnikov showed that if a_{ij} have symmetric distributions and their moments satisfy some mild assumptions, then $Pr[\lambda_1(A) \le 2\sigma\sqrt{n} + o(1)] = 1 - o(1)$. They also stated that a similar result would hold without the symmetric assumption. Furthermore, Soshnikov proved in [So] that under the same assumptions about a_{ij}, the joint distribution of the k dimensional random vector formed by the first k eigenvalues, scaled properly, tends to a weak limit, for any fixed k.

3 Proof of Theorem 1

The key tool of the proof is a powerful concentration result, due to Talagrand [10]. To state this inequality, we first need to define the so-called Talagrand distance in a product space. Let $\Omega_1, \ldots, \Omega_m$ be probability spaces, and let Ω denote their product space. Fix a set $\mathcal{A} \subset \Omega$ and a point $\mathbf{x} = (x_1, \ldots, x_m) \in \Omega$. We say that x has Talagrand distance t from $\mathcal{A}$ if t is the smallest number such that the following holds. For any real vector $\alpha = (\alpha_1, \ldots, \alpha_m)$, there is a point $\mathbf{y} = (y_1, \ldots y_m) \in \mathcal{A}$ such that

$$\sum_{x_i \ne y_i} |\alpha_i| \le t \left(\sum_{i=1}^{n} \alpha_i^2 \right)^{1/2} .$$

Let $\mathcal{A}_t$ denote the set of all points with Talagrand distance at most t from $\mathcal{A}$. Talagrand proved that for any $t \ge 0$,

$$\Pr[\mathcal{A}] Pr[\overline{\mathcal{A}_t}] \le e^{-t^2/4},$$

where $\overline{\mathcal{A}_t}$ denotes the complement of $\mathcal{A}_t$. Remarkably, the rather abstract and difficult definition of the Talagrand distance suits our problem perfectly, as shown in the proof below.

Consider the product space spanned by $a_{ij}, 1 \le i \le j \le n$. A vector in this space corresponds to a random matrix. Let m be a median of λ_1 and let $\mathcal{A}$ be the set of all matrices (vectors) T such that $\lambda_1(A) \le m$. By definition, $Pr[\mathcal{A}] \ge 1/2$. By a well known fact in linear algebra

$$\lambda_1(A) = \max_{\|\mathbf{v}\| = \|\mathbf{w}\| = 1} \sum_{1 \le i < j \le n} (v_i w_j + v_j w_i) t_{ij} + \sum_{i=1}^{n} v_i w_i t_{ii} .$$

Consider a matrix $X = (x_{ij})$ where $\lambda_1(X) \ge m + t$. There are vectors $\mathbf{v} = (v_1, \ldots, v_n)$, $\mathbf{w} = (w_1, \ldots, w_n)$ with norm 1 such that

$$\mathbf{v} X \mathbf{w} = \sum_{1 \le i \le j \le n} (v_i w_j + v_j w_i) x_{ij} + \sum_{i=1}^{n} v_i w_i x_{ii} \ge m + t .$$

On the other hand, for any $Y = (y_{ij}) \in \mathcal{A}$

$$\mathbf{v} Y \mathbf{w} = \sum_{1 \le i \le j \le n} (v_i w_j + v_j w_i) y_{ij} + \sum_{i=1}^{n} v_i w_i y_{ii} \le m .$$

Set $\alpha_{ij} = (v_i w_j + v_j w_i)$ for $1 \le i < j \le n$ and $\alpha_{ii} = v_i w_i$ for $1 \le i \le n$. It is easy to show that

$$\sum_{1 \le i \le j \le n} \alpha_{ij}^2 < 2(\sum_{1 \le i \le n} v_i^2)(\sum_{i=1}^{n} w_i^2) = 2 \ .$$

Since $|x_{ij} - y_{ij}| \le 2$, we have

$$\sum_{x_{ij} \neq y_{ij}} |\alpha_{ij}| \ge t/2 > \frac{t}{\sqrt{8}} \left(\sum_{1 \le i \le j \le n} \alpha_{ij}^2 \right)^{1/2} \ .$$

By definition, it follows that $X \in \overline{\mathcal{A}_{t/\sqrt{8}}}$. Therefore, by Talagrand's inequality

$$Pr[\lambda_1(A) \ge m + t] \le 2e^{-t^2/32} \ . \tag{9}$$

Let $\mathcal{B}$ be the set of A such that $\lambda_1(A) \le m - t$. By a similar argument, one can show that if $\lambda_1(A) \ge m$ then $A \in \overline{\mathcal{B}_{t/\sqrt{8}}}$. Recall that $Pr[\lambda_1(A) \ge m] \ge 1/2$. Thus Talagrand's inequality implies

$$Pr[\lambda_1(A) \le m - t] \le 2e^{-t^2/32} \ . \tag{10}$$

¿From here, one can derive that the difference between the median and the expectation of λ_1 is bounded by a constant:

$$\begin{aligned} |E(\lambda_1(A)) - m| &\le E(|\lambda_1 - m|) \le \int_0^\infty t Pr[|\lambda_1(A) - m| \ge t] dt \\ &\le \int_0^\infty 4te^{-t^2/32} dt = 64 \ . \end{aligned} \tag{11}$$

Inequalities (9), (10) and (11) together imply the desired deviation bound for $\lambda_1(A)$. The statements involving $\delta_1(A)$ and $\delta_n(A)$ can be proved in a similar way, using the following equalities:

$$\delta_1(A) = \max_{\mathbf{x}, \|\mathbf{x}\|=1} \mathbf{x} A \mathbf{x}.$$

$$\delta_n(A) = \min_{\mathbf{x}, \|\mathbf{x}\|=1} \mathbf{x} A \mathbf{x} \ .$$

The sharpness of the result. The following example shows that the bound in Theorem 1 is best possible, up to a multiplicative constant in the exponent.

Assume that a_{ij}, $1 \le i \le j \le n$, have the following distribution: $a_{ij} = 1$ with probability p and $a_{ij} = -p/q$ with probability $q = 1 - p$. A matrix is *fat* if it contains an all 1 principle sub-matrix of size $E[\lambda_1] + t$. It is trivial that if A is fat then $\lambda_1(A) \ge E[\lambda_1] + t$. On the other hand, the probability that a matrix is fat is at least $p^{(E[\lambda_1]+t)^2} = e^{-\left[E(\lambda_1]+t\right)^2 \log \frac{1}{p}}$. Thus, if p is a positive constant and t is of order $\Omega(E[\lambda_1])$, then

$$Pr[|\lambda_1 - E[\lambda_1]| \ge t] \ge e^{-ct^2},$$

for some positive constant c.

4 Proof of Theorem 2

Given a symmetric matrix A, $\lambda_2(A)$ can be expressed as follows [2]:

$$\lambda_2(A) = \min_{\mathbf{0} \neq \mathbf{v} \in R^n} \max_{\substack{\mathbf{x},\mathbf{y} \\ \|\mathbf{x}\|=\|\mathbf{y}\|=1 \\ \mathbf{x}\mathbf{v}=\mathbf{y}\mathbf{v}=0}} \mathbf{x}A\mathbf{y} .$$

Define

$$\mu_2(A) = \max_{\substack{\mathbf{x},\mathbf{y} \\ \mathbf{x}\mathbf{1}=\mathbf{y}\mathbf{1}=0 \\ \|\mathbf{x}\|=\|\mathbf{y}\|=1}} \mathbf{x}A\mathbf{y} .$$

It is clear that $\mu_2(A) \geq \lambda_2(A)$ for any matrix A. In the rest of the proof, we use shorthands μ_2, λ_2 for $\mu_2(A), \lambda_2(A)$, respectively, where A is distributed as described in the theorem formulation. Similar to the previous section, by Talagrand's inequality we can show

Lemma 4.1 *There are positive constants c and K such that for any $t > K$*

$$Pr[|\mu_2 - E(\mu_2)| \geq t] \leq e^{-ct^2} .$$

Set $A' = A - pJ_n$, where J_n denotes the all one matrix of order n. It is easy to show that $\mu_2(A) \leq \lambda_1(A')$. Indeed,

$$\begin{aligned} \mu_2 &= \max_{\substack{\mathbf{x},\mathbf{y} \\ \mathbf{x}\mathbf{1}=\mathbf{y}\mathbf{1}=0 \\ \|bx\|=\|\mathbf{y}\|=1}} \mathbf{x}(A' + pJ_n)\mathbf{y} = \max_{\substack{\mathbf{x},\mathbf{y} \\ \mathbf{x}\mathbf{1}=\mathbf{y}\mathbf{1}=0 \\ \|bx\|=\|\mathbf{y}\|=1}} \mathbf{x}A'\mathbf{y} \\ &\leq \max_{\|bx\|=\|\mathbf{y}\|=1} \mathbf{x}A'\mathbf{y} = \lambda_1(A') , \end{aligned}$$

where the second equality uses the fact that $\mathbf{x}$ and $\mathbf{y}$ are orthogonal to the vector of all 1's and are thus orthogonal to every row of J_n.

Since each non-diagonal entry of A' has mean 0 and is bounded in absolute value by $1 + p \leq 2$, by the result (3) of Füredi and Komlós, $E[\lambda_1(A')] \leq 3\sqrt{n}$. Assume that $t \geq 10\sqrt{n}$; Theorem 1 implies then

$$Pr[|\lambda_2 - E(\lambda_2)| \geq t] \leq Pr[\lambda_1(A') \geq E(\lambda_1(A')) + t/2] \leq e^{-ct^2} .$$

The proof of the case $t < 10\sqrt{n}$ is harder and is based on the following two lemmas.

Lemma 4.2 *For every constant $p > 0$ there exist constants $c_p > 0$, $K_p > 0$ so that for any $K_p < t < 10\sqrt{n}$, there is a positive number $\epsilon_t = O(t^{1/2}(np)^{-1/2})$ such that*

$$Pr[\mu_2 - (1 + \epsilon_t)\lambda_2 \geq t] \leq e^{-c_p t^2} .$$

Lemma 4.3 *For every constant $p > 0$ there exists a constant $L_p > 0$ such that*

$$E[\mu_2] - E[\lambda_2] \leq L_p .$$

Assuming these two lemmas hold, we can finish the proof as follows. First assume, without loss of generality, that $t \geq 5L_p$. Consider the upper tail:

$$\begin{aligned} Pr[\lambda_2 \geq E(\lambda_2) + t] &\leq Pr[\mu_2 \geq E(\lambda_2) + t] \\ &\leq Pr[\mu_2 \geq E(\mu_2) + (t - L_p)] \\ &\leq e^{-\Omega((t-L_p)^2)} = e^{-c_p t^2} , \end{aligned}$$

by Lemma 4.1.

Now consider the lower tail:

$$\begin{aligned} Pr[\lambda_2 \leq E[\lambda_2] - t] &\leq Pr[(1+\epsilon_t)\lambda_2 \leq (1+\epsilon_t)E[\lambda_2] - t] \\ &\leq Pr[\mu_2 \leq (1+\epsilon_t)E[\lambda_2] - t/2] + Pr[\mu_2 - (1+\epsilon_t)\lambda_2 \geq t/2] . \end{aligned}$$

By Lemma 4.2

$$Pr[\mu_2 - (1+\epsilon_t)\lambda_2 \geq t/2] \leq e^{-c_p t^2} .$$

On the other hand,

$$Pr[\mu_2 \leq (1+\epsilon_t)E[\lambda_2] - t/2] \leq \Pr[\mu_2 \leq (1+\epsilon_t)E[\mu_2] - t/2] .$$

Given that t is sufficiently large, $\epsilon_t E[\mu_t] = O(t^{1/2}) \leq t/4$. So, by Lemma 4.1, the last probability can also be bounded by $e^{-c_p t^2}$ and this completes the proof. □

To prove Lemmas 4.2 and 4.3 we need three other lemmas. The first two (Lemmas 4.4 and 4.5) are linear algebraic statements. The last one (Lemma 4.6) is a statement about the concentration of a certain random variable, which is a function of the entries a_{ij} of the random symmetric matrix A.

Lemma 4.4 *Let A be an n by n real symmetric matrix. Let a satisfy $0 \leq a < \sqrt{n}$. Denote by $\mathbf{v}_1$ a unit eigenvector corresponding to $\lambda_1(A)$. Assume there is a number $c_1, 0 < |c_1| \leq \sqrt{n}$ such that $\|\mathbf{1} - c_1\mathbf{v}_1\| \leq a$. Then*

$$\mu_2(A) - \lambda_2(A) \leq \frac{2a\lambda_2(A)}{\sqrt{n} - a} + \frac{a^2\lambda_1(A)}{(\sqrt{n} - a)^2} .$$

Proof. Note first that

$$\|c_1\mathbf{v}_1\| = \|(c_1\mathbf{v}_1 - \mathbf{1}) + \mathbf{1}\| \geq \|\mathbf{1}\| - \|c_1\mathbf{v}_1 - 1\| \geq \sqrt{n} - a .$$

Assume that $\mu_2(A) = \mathbf{x}A\mathbf{y}$, where $\mathbf{x}, \mathbf{y}$ are unit vectors perpendicular to $\mathbf{1}$. Then

$$\mathbf{x}(c_1\mathbf{v}_1) = \mathbf{x}, (c_1\mathbf{v}_1 - \mathbf{1} + \mathbf{1}) = \mathbf{x}(c_1\mathbf{v}_1 - 1) \leq \|\mathbf{x}\| \cdot \|c_1\mathbf{v}_1 - \mathbf{1}\| \leq a .$$

Notice that as $\|\mathbf{1}\| = \sqrt{n}y$ and $a < \sqrt{n}$, we have $c_1 \neq 0$. Define

$$\mathbf{x}' = x - \frac{\mathbf{x}(c_1\mathbf{v}_1)}{(c_1\mathbf{v}_1)(c_1\mathbf{v}_1)} c_1\mathbf{v}_1 .$$

Then $\mathbf{x}'$ is orthogonal to $\mathbf{1}$ and satisfies $\|\mathbf{x}'\| \leq \|\mathbf{x}\| = 1$. Set $\mathbf{u} = \mathbf{x} - \mathbf{x}'$. Then

$$\|\mathbf{u}\| = \frac{|\mathbf{x}(c_1\mathbf{v}_1)|}{\|c_1\mathbf{v}_1\|} \leq \frac{a}{\sqrt{n} - a} \ .$$

Similarly, set

$$\mathbf{y}' = \mathbf{y} - \frac{\mathbf{y}(c_1\mathbf{v}_1)}{(c_1\mathbf{v}_1)(c_1\mathbf{v}_1)} c_1\mathbf{v}_1 \ ,$$

then $\mathbf{y}'$ is a vector of norm at most 1 orthogonal to $\mathbf{v}_1$. Denoting $\mathbf{w} = \mathbf{y} - \mathbf{y}'$, we can prove that $\|\mathbf{w}\| \leq a/(\sqrt{n} - a)$.

By definition, $\lambda_2(A) \geq |\mathbf{x}'A\mathbf{y}'|$. On the other hand, by the Cauchy–Schwartz inequality

$$\begin{aligned}
\mathbf{x}A\mathbf{y} - \mathbf{x}'A\mathbf{y}' &= (\mathbf{x}' + \mathbf{u})A(\mathbf{y}' + \mathbf{w}) - \mathbf{x}'A\mathbf{y}' = \mathbf{w}A\mathbf{x}' + \mathbf{u}A\mathbf{y}' + \mathbf{u}A\mathbf{w} \\
&\leq \|\mathbf{w}\|\|A\mathbf{x}'\| + \|u\|\|A\mathbf{y}'\| + \lambda_1(A)\|u\|\|w\| \ .
\end{aligned}$$

Recall that $\mathbf{x}', \mathbf{y}'$ are orthogonal to the first eigenvector of A. Therefore, $\|\mathbf{x}'A\|$ and $\|A\mathbf{y}'\|$ are at most $\lambda_2(A)$. Then

$$\mu_2(a) - \lambda_2(A) \leq \frac{a\lambda_2(A)}{\sqrt{n} - a} + \frac{a\lambda_2(A)}{\sqrt{n} - a} + \frac{a^2\lambda_1(A)}{(\sqrt{n} - a)^2} = \frac{2a\lambda_2(A)}{\sqrt{n} - a} + \frac{a^2\lambda_1(A)}{(\sqrt{n} - a)^2} \ ,$$

and the lemma follows. □

Lemma 4.5 *Let $A = (a_{ij})$ be an n by n real symmetric matrix. Assume that s and X are positive numbers satisfying $\lambda_2(A) \leq s/2$ and $\sum_{i=1}^n (\sum_{j=1}^n a_{ij} - s)^2 \leq X$. Then there is a number $c_1, |c_1| \leq \sqrt{n}$ such that $\|\mathbf{1} - c_1\mathbf{v}_1\| \leq 2\sqrt{X}/s$, where $\mathbf{v}_1$ is a unit eigenvector corresponding to $\lambda_1(A)$.*

Proof. . Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be unit eigenvectors of A, corresponding to the eigenvalues $\lambda_1(A), \ldots, \lambda_n(A)$, respectively. Since these vectors form an orthogonal basis of R^n, we can express the vector $\mathbf{1}$ as their linear combination:

$$\mathbf{1} = \sum_{i=1}^n c_i\mathbf{v}_i \ ,$$

where $|c_1| \leq \|\mathbf{1}\| = \sqrt{n}$. It is not too difficult to check the following relations:

$$\sum_{i=1}^n (\lambda_i(A) - s)^2 c_i^2 = \|(A - sI)\mathbf{1}\|^2 = \sum_{i=1}^n (\sum_{j=1}^n a_{ij} - s)^2 \ .$$

By the assumptions of the lemma we get then

$$X \geq \sum_{i=1}^n (\sum_{j=1}^n a_{ij} - s)^2 \geq \sum_{i=2}^n (\lambda_i(A) - s)^2 c_i^2 \geq \frac{s^2}{4} \sum_{i=2}^n c_i^2 \ .$$

Therefore,

$$\|\mathbf{1} - c_1\mathbf{v}_1\|^2 = \sum_{i=2}^n c_i^2 \leq \frac{4X}{s^2} \ ,$$

as desired. □

Lemma 4.6 *Let a_{ij}, $1 \le j \le i \le n$ be independent random variables, uniformly bounded by 1 in their absolute values. Assume that for $i > j$, the a_{ij} have a common expectation p. Define $a_{ij} = a_{ji}$ for $j > i$. Then there exists an absolute constant $c > 0$ so that for all $t > 1$,*

$$Pr\left[\sum_{i=1}\left(\sum_{j=1} a_{ij} - np\right)^2 \ge tn^2\right] < e^{-ct^2} \ .$$

Proof. For $1 \le i \le n$, let $p_i = E[a_{ii}]$. We define

$$Y_i = (\sum_{j=1}^{n} a_{ij} - np)^2 \,,$$

then $Y = \sum_{i=1}^{n}(\sum_{j=1}^{n} a_{ij} - np)^2 = \sum_{i=1}^{n} Y_i$.

We first estimate from above the expectation of Y_i. Set $b_{ij} = a_{ij}$ for all $j \ne i$, set also $b_{ii} = a_{ii} + p - p_i$. Then $E[b_{ij}] = p$ for all $1 \le i, j \le n$. We obtain:

$$Y_i = \left(\sum_{j=1}^{n}(b_{ij} - np) + (p_i - p)i\right)^2 = (\sum_{j=1}^{n} b_{ij} - np)^2 + 2(p_i - p)\sum_{j=1}^{n} b_{ij} - np) + (p_i - p)^2 \ .$$

Recall that b_{ij} are independent random variables with a common mean p. Therefore

$$\begin{aligned} E[Y_i] &= E[\sum_{j=1}^{n} b_{ij} - np)^2] + (p_i - p)^2 \\ &= VAR[\sum_{j=1}^{n} b_{ij}] + (p_i - p)^2 = \sum_{j=1}^{n} VAR[b_{ij}] + (p_i - p)^2 \le n(1-p) + (p_i - p)^2 \\ &\le n \,, \end{aligned}$$

for large enough n. This implies that $E[Y] = \sum_{i=1}^{n} E[Y_i] \le n^2$.

Now, it is easy to see that for every $1 \le j \le i \le n$, changing the value of the random variable a_{ij} can change the value of Y by at most $c_{ij} = O(n)$ (recall the assumption $|a_{ij}| \le 1$). Then the so called "independent bounded difference inequality", proved by applying the Azuma–Hoeffding martingale inequality (see,. e.g., [7]), asserts that for every $h > 0$,

$$Pr[Y - E[Y] \ge h] \le \exp\{-h^2/2 \sum_{1 \le j \le i \le n} c_{ij}^2\} \le \exp\{-h^2/O(n^4)\} \ .$$

Substituting $h = (t-1)n^2$ and using the fact $E[Y] \le n^2$, we get the desired bound on the upper tail of Y. □

Proof of Lemma 4.2. Recall that by (2 we have $E[\lambda_1[A] = O(\sqrt{n})$. From the analysis of the case $t \ge 10\sqrt{n}$ it follows then that $Pr[\lambda_2 \ge np/2] \le e^{-c(np)^2} \le e^{-ct^2}$. Also, by Corollary 2.1 $E[\lambda_1(A)] = np + o(n)$. Combined with our Theorem 1, this implies that $Pr[\lambda_1(A) \ge 2np] \le e^{-c(np)^2} \le e^{-ct^2}$. These two facts, together with Lemma 4.6, show that for that if $t < 10\sqrt{n}$, then with probability at least $1 - e^{-ct^2}$, the following three properties hold:

1. $\sum_{i=1}^{n}(\sum_{j=1}^{n} a_{ij} - np)^2 \leq n^2 t$;
2. $\lambda_2(A) \leq np/2$;
3. $\lambda_1(A) \leq 2np$.

Assume that a matrix A satisfies conditions 1, 2 and 3 above. Applying (in this order) Lemma 4.5 with $X = n^2 t$ and $s = np$ and Lemma 4.4 with $a = 2X^{1/2}/s = 2t^{1/2}/p$, we have that with probability at least $1 - e^{-ct^2}$

$$\mu_2(A) - \left(1 + \frac{2a}{\sqrt{n} - a}\right)\lambda_2(A) \leq \frac{a^2 \lambda_1(A)}{(\sqrt{n} - a)^2} \leq \frac{4X}{s^2}\, \frac{2np}{(\sqrt{n} - 2\sqrt{X}/s)^2} \leq \frac{9t}{p} \ .$$

Substituting the value of a, we get:

$$Pr[\mu_2 - \left(1 + \frac{5\sqrt{t}}{\sqrt{np}}\right)\lambda_2 \geq \frac{9t}{p}] < e^{-ct^2} \ .$$

The proof is completed by rescaling, namely, by setting $t := t/p$. $\square$

Proof of Lemma 4.3. First notice that

$$E[\mu_2] - E[\lambda_2] = E[\mu_2 - \lambda_2] \leq \int_0^\infty tPr[\mu_2 - \lambda_2 \geq t]dt \ .$$

Moreover,

$$\int_0^\infty tPr[\mu_2 - \lambda_2 \geq t]dt \leq \int_0^{K_p} tdt + \int_{K_p}^{10\sqrt{n}} tPr[\mu_2 - \lambda_2 \geq t]dt + \int_{10\sqrt{n}}^\infty tPr[\mu_2 \geq t]dt \ .$$

The first integral is clearly bounded by a constant depending on pr only. By Lemma 4.1 and the fact that $E(\mu_2) \leq 3\sqrt{n}$, $Pr[\mu_2 \geq t] \leq e^{-ct^2}$ for $t \geq 10\sqrt{n}$. Thus $\int_{10\sqrt{n}}^\infty tPr[\mu_2 \geq t]dt \leq \int_0^\infty te^{-ct^2}dt = O(1)$. To bound the second integral, note that

$$\begin{aligned}\int_{K_p}^{10\sqrt{n}} tPr[\mu_2 - \lambda_2 \geq t]dt \ &\leq \ \int_{K_p}^{10\sqrt{n}} tPr[\mu_2 - \lambda_2 \geq t/2 + \epsilon_t \lambda_2]dt \\ &+ \ \int_{K_p}^{10\sqrt{n}} tPr[\epsilon_t \lambda_2 \geq t/2]dt.\end{aligned}$$

By Lemma 4.2,

$$\int_{K_p}^{10\sqrt{n}} tPr[\mu_2 - \lambda_2 \geq t/2 + \epsilon_t \lambda_2]dt \leq \int_0^{10\sqrt{n}} te^{-c_p t^2}dt = l_1 \ ,$$

where $l_1 > 0$ is a constant depending only on p.

On the other hand, we know that $\epsilon_t \leq bt^{1/2}(np)^{-1/2}$ for some constant b. Using the analysis of the case $t \geq 10\sqrt{n}$, assume that $K_p > (30b/p)^2$; for any $K \leq t \leq 10\sqrt{n}$ we have:

$$Pr[\epsilon_t \lambda_2 \geq t/2] \leq Pr[\lambda_2 \geq \frac{t^{1/2}}{2b}(np)^{1/2}] \leq e^{-c_p t^2} \ .$$

This implies that

$$\int_{K_p}^{10\sqrt{n}} tPr[\epsilon_t\lambda_2 \geq t/2]dt \leq \int_0^{10\sqrt{n}} te^{-c_p t^2}dt = l_2 \ ,$$

where l_2 is a constant depending on p only. This completes the proof. □

The proof for δ_2 is similar. Instead of μ_2, consider $\mu_2' = \max_{\mathbf{x},\|\mathbf{x}=1\|,\mathbf{x}\mathbf{1}=0} \mathbf{x}A\mathbf{x}$. Again, using Talagrand's inequality one can obtain a version of Lemma 4.1 for μ_2'. The rest of the proof is similar and we omit the details.

5 Concluding remarks

- Unfortunately, we are unable to extend Theorem 2 to the case when the expectation p of the non-diagonal entries of a random matrix A_n is a function of n and tends to zero as n tends to infinity, without imposing additional restrictions of the distribution of entries. However, Theorem 2 can be extended in the following special but important case: the diagonal entries of $A_n = (a_{ij})$ are all zeroes, and the entries above the main diagonal are i.i.d. Bernoulli random variables with parameter $p = p(n)$, i.e., $Pr[a_{ij} = 1] = p$ and $Pr[a_{ij} = 0] = 1 - p$ for all $1 \leq i < j \leq n$. In this case the random matrix A_n can be identified with the adjacency matrix of a *random graph* $G(n,p)$, and the eigenvalues of A_n are the eigenvalues of a random graph on n vertices. Under these assumptions we have the following result.

  Theorem 3 *There are positive constants c and K such that if $p = \omega(n^{-1})$ then for any $t > K$,*

  $$Pr[|\lambda_2(A) - E[\lambda_2(A)]| \geq t] \leq e^{-ct^2},$$

  where A is the adjacency matrix of $G(n,p)$. The same result holds for $\delta_2(A)$.

  This theorem can be proved by repeating the arguments in the proof of Theorem 2 under the new assumptions. We have to make some significant changes only in the proof of Lemma 4.6. The method of bounded difference martingale (Azuma-Hoeffding's inequality) seems not powerful enough to prove the the statement Lemma 4.6 when p is decreasing in n, and we need to invoke a recent concentration technique presented in [11]. The details are omitted. Notice that in many graph theoretic applications the eigenvalue $\lambda_2(A(G))$ is of special importance as it reflects such graph properties as expansion, convergence of a random walk to the stationary distribution etc.

- Though we could show the tightness of our main result (Theorem 1) in some cases and for some values of the deviation parameter t, it will be extremely interesting to reach a deeper understanding of the tightness of Theorem 1 for the whole range of t and for some particular important distributions of the entries of A.

- Theorem 1 is obtained under very general assumptions on the distribution of the entries of a symmetric matrix A. Still, it will be very desirable to generalize our result even further, in particular, dropping or weakening the restrictive assumption about the uniform boundness of the entries of A. This task however may require completely different tools as the Talagrand inequality appears to be suited for the case of bounded random variables.

- Finally, it would be quite interesting to find further applications of our concentration results in algorithmic problems on graphs. The ability to compute the eigenvalues of a graph in polynomial time combined with an understanding of potentially rich structural information encoded by the eigenvalues can certainly provide a basis for new algorithmic results exploiting eigenvalues of graphs and their concentration.

Acknowledgment. The authors are grateful to Zeev Rudnick for his helpful comments.

References

[1] L. Arnold, *On Wigner semi-circle law for the eigenvalues of random matrices*, Z. Wahrscheinlichkeitstheorie Verw. Gebiete, 19, 191–198 (1971).

[2] F. R. Gantmacher, **Applications of the theory of matrices**, Intersciences, New York, 1959.

[3] Z. Füredi and J. Komlós, *The eigenvalues of random symmetric matrices*, Combinatorica 1 (3), 233–241 (1981).

[4] F. Juhász, *On the spectrum of a random graph*, in: Algebraic method in graph theory (L. Lovász et al, eds.), Coll. Math. Soc. J. Bolyai 25, North Holland, 313–316, 1981.

[5] M. Krivelevich and V. H. Vu, *Approximating the independence number and the chromatics number in expected polynomial time*, Proceedings of the 7^{th} Int. Colloq. on Automata, Languages and Programming (ICALP'2000), 13-25.

[6] M. L. Mehta, **Random matrices**, Academic Press, New York, 1991.

[7] C. J. H. McDiarmid, On the method of bounded differences, in *Surveys in Combinatorics 1989*, London Math. Soc. Lecture Notes Series 141 (Siemons J., ed.), Cambridge Univ.

[8] Ya. G. Sinai and A.B. Soshnikov, *A refinement of Wigner's semi-circle law in a neighborhood of the spectrum edge for random symmetric matrices*, Functional Anal. and its appl., Vol 32 (2), 114–131 (1998).

[9] A. Soshnikov, *Universality of edge of the spectrum in Wigner random matrices*, manuscript, http://front.math.ucdavis.edu/math-ph/9907013.

[10] M. Talagrand, *Concentration of Measures and Isoperimetric Inequalities in product spaces*, Publications Mathematiques de l'I.H.E.S., 81, 73-205 (1996).

[11] V. H. Vu. *A large concentration result on the number of subgraphs in a random graph*, Combinatorics, Probability and Computing, to appear.

[12] E. Wigner, *On the distribution of the roots of certain symmetric matrices,* Ann. Math. 67, 325–328 (1958).