Sparse random graphs: Eigenvalues and Eigenvectors

Linh V. Tran, Van H. Vu* and Ke Wang
Department of Mathematics, Rutgers, Piscataway, NJ 08854

Abstract

In this paper we prove the semi-circular law for the eigenvalues of regular random graph $G_{n,d}$ in the case $d \to \infty$, complementing a previous result of McKay for fixed d. We also obtain a upper bound on the infinity norm of eigenvectors of Erdős-Rényi random graph $G(n,p)$, answering a question raised by Dekel-Lee-Linial.

1 Introduction

1.1 Overview

In this paper, we consider two models of random graphs, the Erdős-Rényi random graph $G(n,p)$ and the random regular graph $G_{n,d}$. Given a real number $p = p(n)$,$0 \leq p \leq 1$, the Erdős-Rényi graph on a vertex set of size n is obtained by drawing an edge between each pair of vertices, randomly and independently, with probability p. On the other hand, $G_{n,d}$, where $d = d(n)$ denotes the degree, is a random graph chosen uniformly from the set of all simple d-regular graphs on n vertices. These are basic models in the theory of random graphs. For further information, we refer the readers to the excellent monographs [4], [19] and survey [33].

Given a graph G on n vertices, the adjacency matrix A of G is an $n \times n$ matrix whose entry a_{ij} equals one if there is an edge between the vertices i and j and zero otherwise. All diagonal entries a_{ii} are defined to be zero. The eigenvalues and eigenvectors of A carry valuable information about the structure of the graph and have been studied by many researchers for quite some time, with both theoretical and practical motivations (see, for example, [2], [3], [12], [25] [16], [13], [15], [14], [30], [10], [27], [24]).

The goal of this paper is to study the eigenvalues and eigenvectors of $G(n,p)$ and $G_{n,d}$. We are going to consider:

*V. Vu is supported by NSF grants DMS-0901216 and AFOSAR-FA-9550-09-1-0167.

- The global law for the limit of the empirical spectral distribution (ESD) of adjacency matrices of $G(n,p)$ and $G_{n,d}$. For $p=\omega(1/n)$, it is well-known that eigenvalues of $G(n,p)$ (after a proper scaling) follows Wigner's semicircle law (we include a short proof in the Appendix A for completeness). Our main new result shows that the same law holds for random regular graph with $d \to \infty$ with n. This complements the well known result of McKay for the case when d is an absolute constant (McKay's law) and extends recent results of Dumitriu and Pal [9] (see Section 1.2 for more discussion).

- Bound on the infinity norm of the eigenvectors. We first prove that the infinity norm of any (unit) eigenvector v of $G(n,p)$ is almost surely $o(1)$ for $p=\omega(\log n/n)$. This gives a positive answer to a question raised by Dekel, Lee and Linial [7]. Furthermore, we can show that v satisfies the bound $\|v\|_\infty = O\left(\sqrt{\log^{2.2} g(n) \log n/np}\right)$ for $p=\omega(\log n/n)=g(n)\log n/n$, as long as the corresponding eigenvalue is bounded away from the (normalized) extremal values -2 and 2.

We finish this section with some notation and conventions.

Given an $n\times n$ symmetric matrix M, we denote its n eigenvalues as

$$\lambda_1(M) \le \lambda_2(M) \le \ldots \le \lambda_n(M),$$

and let $u_1(M),\ldots,u_n(M) \in \mathbb{R}^n$ be an orthonormal basis of eigenvectors of M with

$$Mu_i(M) = \lambda_i u_i(M).$$

The empirical spectral distribution (ESD) of the matrix M is a one-dimensional function

$$F_n^{\mathbf{M}}(x) = \frac{1}{n}|\{1 \le j \le n : \lambda_j(M) \le x\}|,$$

where we use $|\mathbf{I}|$ to denote the cardinality of a set $\mathbf{I}$.

Let A_n be the adjacency matrix of $G(n,p)$. Thus A_n is a random symmetric $n\times n$ matrix whose upper triangular entries are iid copies of a real random variable ξ and diagonal entries are 0. ξ is a Bernoulli random variable that takes values 1 with probability p and 0 with probability $1-p$.

$$\mathbb{E}\xi = p, \mathbb{V}ar\xi = p(1-p) = \sigma^2.$$

Usually it is more convenient to study the normalized matrix

$$M_n = \frac{1}{\sigma}(A_n - pJ_n)$$

where J_n is the $n \times n$ matrix all of whose entries are 1. M_n has entries with mean zero and variance one. The global properties of the eigenvalues of A_n and M_n are essentially the same (after proper scaling), thanks to the following lemma

Lemma 1.1. *(Lemma 36, [30]) Let A, B be symmetric matrices of the same size where B has rank one. Then for any interval I,*

$$|N_I(A+B) - N_I(A)| \leq 1,$$

where $N_I(M)$ is the number of eigenvalues of M in I.

Definition 1.2. *Let E be an event depending on n. Then E holds with overwhelming probability if $\mathbf{P}(E) \geq 1 - \exp(-\omega(\log n))$.*

The main advantage of this definition is that if we have a polynomial number of events, each of which holds with overwhelming probability, then their intersection also holds with overwhelming probability.

Asymptotic notation is used under the assumption that $n \to \infty$. For functions f and g of parameter n, we use the following notation as $n \to \infty$: $f = O(g)$ if $|f|/|g|$ is bounded from above; $f = o(g)$ if $f/g \to 0$; $f = \omega(g)$ if $|f|/|g| \to \infty$, or equivalently, $g = o(f)$; $f = \Omega(g)$ if $g = O(f)$; $f = \Theta(g)$ if $f = O(g)$ and $g = O(f)$.

1.2 The semicircle law

In 1950s, Wigner [32] discovered the famous semi-circle for the limiting distribution of the eigenvalues of random matrices. His proof extends, without difficulty, to the adjacency matrix of $G(n,p)$, given that $np \to \infty$ with n. (See Figure 1 for a numerical simulation)

Theorem 1.3. *For $p = \omega(\frac{1}{n})$, the empirical spectral distribution (ESD) of the matrix $\frac{1}{\sqrt{n}\sigma}A_n$ converges in distribution to the semicircle distribution which has a density $\rho_{sc}(x)$ with support on $[-2,2]$,*

$$\rho_{sc}(x) := \frac{1}{2\pi}\sqrt{4-x^2}.$$

If $np = O(1)$, the semicircle law no longer holds. In this case, the graph almost surely has $\Theta(n)$ isolated vertices, so in the limiting distribution, the point 0 will have positive constant mass.

The case of random regular graph, $G_{n,d}$, was considered by McKay [21] about 30 years ago. He proved that if d is fixed, and $n \to \infty$, then the limiting density function is

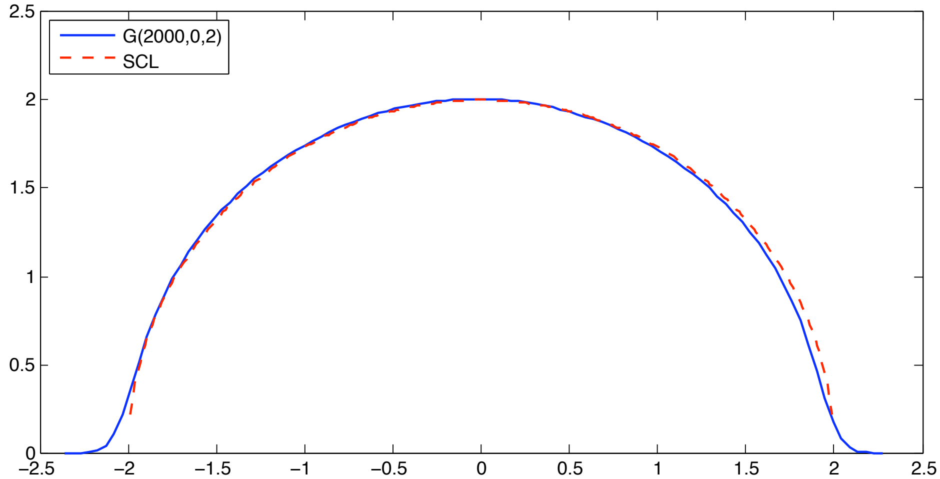

Figure 1: The probability density function of the ESD of $G(2000, 0.2)$

$$f_d(x) = \begin{cases} \frac{d\sqrt{4(d-1)-x^2}}{2\pi(d^2-x^2)}, & \text{if } |x| \leq 2\sqrt{d-1}; \\ \\ 0 & \text{otherwise.} \end{cases}$$

This is usually referred to as McKay or Kesten-McKay law.

It is easy to verify that as $d \to \infty$, if we normalize the variable x by $\sqrt{d-1}$, then the above density converges to the semicircle distribution on $[-2, 2]$. In fact, a numerical simulation shows the convergence is quite fast(see Figure 2).

It is thus natural to conjecture that Theorem 1.3 holds for $G_{n,d}$ with $d \to \infty$. Let A'_n be the adjacency matrix of $G_{n,d}$, and set

$$M'_n = \frac{1}{\sqrt{\frac{d}{n}(1 - \frac{d}{n})}}(A'_n - \frac{d}{n}J).$$

Conjecture 1.4. *If $d \to \infty$ then the ESD of $\frac{1}{\sqrt{n}}M'_n$ converges to the standard semicircle distribution.*

Nothing has been proved about this conjecture, until recently. In [9], Dimitriu and Pal showed that the conjecture holds for d tending to infinity slowly, $d = n^{o(1)}$. Their method does not extend to larger d.

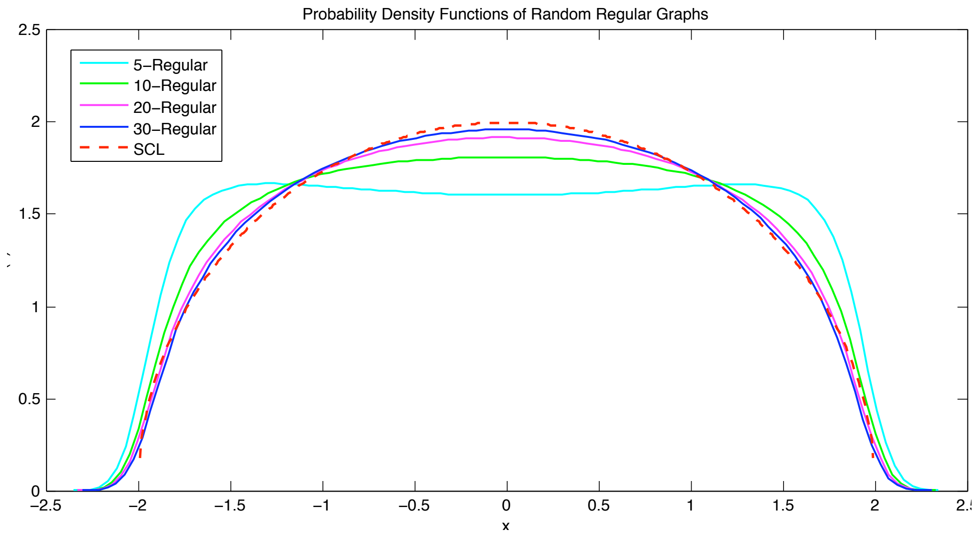

Figure 2: The probability density function of the ESD of Random d-regular graphs with 1000 vertices

We are going to establish Conjecture 1.4 in full generality. Our method is very different from that of [9].

Without loss of generality we may assume $d \leq n/2$, since the adjacency matrix of the complement graph of $G_{n,d}$ may be written as $J_n - A'_n$, thus by Lemma 1.1 will have the spectrum interlacing between the set $\{-\lambda_n(A'_n), \dots, -\lambda_1(A'_n)\}$. Since the semi-circular distribution is symmetric, the ESD of $G_{n,d}$ will converges to semi-circular law if and only if the ESD of its complement does.

Theorem 1.5. *If d tends to infinity with n, then the empirical spectral distribution of $\frac{1}{\sqrt{n}} M'_n$ converges in distribution to the semicircle distribution.*

Theorem 1.5 is a direct consequence of the following stronger result, which shows convergence at small scales. For an interval I let N'_I be the number of eigenvalues of M'_n in I.

Theorem 1.6. *(Concentration for ESD of $G_{n,d}$). Let $\delta > 0$ and consider the model $G_{n,d}$. If d tends to ∞ as $n \to \infty$ then for any interval $I \subset [-2, 2]$ with length at least $\delta^{-4/5} d^{-1/10} \log^{1/5} d$, we have*

$$|N'_I - n \int_I \rho_{sc}(x)dx| < \delta n \int_I \rho_{sc}(x)dx$$

with probability at least $1 - O(\exp(-cn\sqrt{d}\log d))$.

Remark 1.7. *Theorem 1.6 implies that with probability $1 - o(1)$, for $d = n^{\Theta(1)}$, the rank of $G_{n,d}$ is at least $n - n^c$ for some constant $0 < c < 1$ (which can be computed explicitly from the* *lemmas). This is a partial result toward the conjecture by the second author that $G_{n,d}$ almost surely has full rank (see [31]).*

1.3 Infinity norm of the eigenvectors

Relatively little is known for eigenvectors in both random graph models under study. In [7], Dekel, Lee and Linial, motivated by the study of nodal domains, raised the following question.

Question 1.8. *Is it true that almost surely every eigenvector u of $G(n,p)$ has $||u||_\infty = o(1)$?*

Later, in their journal paper [8], the authors added one sharper question.

Question 1.9. *Is it true that almost surely every eigenvector u of $G(n,p)$ has $||u||_\infty = n^{-1/2+o(1)}$?*

The bound $n^{-1/2+o(1)}$ was also conjectured by the second author of this paper in an NSF proposal (submitted Oct 2008). He and Tao [30] proved this bound for eigenvectors corresponding to the eigenvalues in the bulk of the spectrum for the case $p = 1/2$. If one defines the adjacency matrix by writting -1 for non-edges, then this bound holds for all eigenvectors [30, 29].

The above two questions were raised under the assumption that p is a constant in the interval $(0,1)$. For p depending on n, the statements may fail. If $p \le \frac{(1-\epsilon)\log n}{n}$, then the graph has (with high probability) isolated vertices and so one cannot expect that $\|u\|_\infty = o(1)$ for every eigenvector u. We raise the following questions:

Question 1.10. *Assume $p \ge \frac{(1+\epsilon)\log n}{n}$ for some constant $\epsilon > 0$. Is it true that almost surely every eigenvector u of $G(n,p)$ has $||u||_\infty = o(1)$?*

Question 1.11. *Assume $p \ge \frac{(1+\epsilon)\log n}{n}$ for some constant $\epsilon > 0$. Is it true that almost surely every eigenvector u of $G(n,p)$ has $||u||_\infty = n^{-1/2+o(1)}$?*

Similarly, we can ask the above questions for $G_{n,d}$:

Question 1.12. *Assume $d \ge (1+\epsilon)\log n$ for some constant $\epsilon > 0$. Is it true that almost surely every eigenvector u of $G_{n,d}$ has $||u||_\infty = o(1)$?*

Question 1.13. *Assume $d \ge (1+\epsilon)\log n$ for some constant $\epsilon > 0$. Is it true that almost surely every eigenvector u of $G_{n,d}$ has $||u||_\infty = n^{-1/2+o(1)}$?*

As far as random regular graphs is concerned, Dumitriu and Pal [9] and Brook and Lindenstrauss [5] showed that for any normalized eigenvector of a sparse random regular graph is delocalized in the sense that one can not have too much mass on a small set of coordinates. The readers may want to consult their papers for explicit statements.

We generalize our questions by the following conjectures:

Conjecture 1.14. *Assume $p \geq \frac{(1+\epsilon)\log n}{n}$ for some constant $\epsilon > 0$. Let v be a random unit vector whose distribution is uniform in the $(n-1)$-dimensional unit sphere. Let u be a unit eigenvector of $G(n,p)$ and w be any fixed n-dimensional vector. Then for any $\delta > 0$*

$$\mathbf{P}(|w \cdot u - w \cdot v| > \delta) = o(1).$$

Conjecture 1.15. *Assume $d \geq (1+\epsilon)\log n$ for some constant $\epsilon > 0$. Let v be a random unit vector whose distribution is uniform in the $(n-1)$-dimensional unit sphere. Let u be a unit eigenvector of $G_{n,d}$ and w be any fixed n-dimensional vector. Then for any $\delta > 0$*

$$\mathbf{P}(|w \cdot u - w \cdot v| > \delta) = o(1).$$

In this paper, we focus on $G(n,p)$. Our main result settles (positively) Question 1.8 and almost Question 1.10 . This result follows from Corollary 2.3 obtained in Section 2.

Theorem 1.16. *(Infinity norm of eigenvectors) Let $p = \omega(\log n/n)$ and let A_n be the adjacency matrix of $G(n,p)$. Then there exists an orthonormal basis of eigenvectors of A_n, $\{u_1, \dots, u_n\}$, such that for every $1 \leq i \leq n$, $||u_i||_\infty = o(1)$ almost surely.*

For Questions 1.9 and 1.11, we obtain a good quantitative bound for those eigenvectors which correspond to eigenvalues bounded away from the edge of the spectrum.

For convenience, in the case when $p = \omega(\log n/n) \in (0,1)$, we write

$$p = \frac{g(n)\log n}{n},$$

where $g(n)$ is a positive function such that $g(n) \to \infty$ as $n \to \infty$ ($g(n)$ can tend to ∞ arbitrarily slowly).

Theorem 1.17. *Assume $p = g(n)\log n/n \in (0,1)$, where $g(n)$ is defined as above. Let $B_n = \frac{1}{\sqrt{n}\sigma}A_n$. For any $\kappa > 0$, and any $1 \leq i \leq n$ with $\lambda_i(B_n) \in [-2+\kappa, 2-\kappa]$, there exists a corresponding eigenvector u_i such that $||u_i||_\infty = O_\kappa(\sqrt{\frac{\log^{2.2} g(n) \log n}{np}})$with overwhelming probability.*

The proofs are adaptations of a recent approach developed in random matrix theory (as in [30],[29],[10], [11]). The main technical lemma is a concentration theorem about the number of eigenvalues on a finer scale for $p = \omega(\log n/n)$.

2 Semicircle law for regular random graphs

2.1 Proof of Theorem 1.6

We use the method of comparison. An important lemma is the following

Lemma 2.1. *If $np \to \infty$ then $G(n,p)$ is np-regular with probability at least* $\exp(-O(n(np)^{1/2})$.

For the range $p \geq \log^2 n/n$, Lemma 2.1 is a consequence of a result of Shamir and Upfal [26] (see also [20]). For smaller values of np, McKay and Wormald [23] calculated precisely the probability that $G(n,p)$ is np-regular, using the fact that the joint distribution of the degree sequence of $G(n,p)$ can be approximated by a simple model derived from independent random variables with binomial distribution. Alternatively, one may calculate the same probability directly using the asymptotic formula for the number of d-regular graphs on n vertices (again by McKay and Wormald [22]). Either way, for $p = o(1/\sqrt{n})$, we know that

$$\mathbf{P}(G(n,p) \text{ is } np\text{-regular}) \geq \Theta(\exp(-n\log(\sqrt{np})).$$

which is better than claimed in Lemma 2.1.

Another key ingredient is the following concentration lemma, which may be of independent interest.

Lemma 2.2. *Let M be a $n \times n$ Hermitian random matrix whose off-diagonal entries ξ_{ij} are i.i.d. random variables with mean zero, variance 1 and $|\xi_{ij}| < K$ for some common constant K. Fix $\delta > 0$ and assume that the forth moment $M_4 := \sup_{i,j} \mathbf{E}(|\omega_{ij}|^4) = o(n)$. Then for any interval $I \subset [-2,2]$ whose length is at least $\Omega(\delta^{-2/3}(M_4/n)^{1/3})$, the number N_I of the eigenvalues of $\frac{1}{\sqrt{n}}M$ which belong to I satisfies the following concentration inequality*

$$\mathbf{P}(|N_I - n\int_I \rho_{sc}(t)dt| > \delta n \int_I \rho_{sc}(t)dt) \leq 4\exp(-c\frac{\delta^4 n^2 |I|^5}{K^2}).$$

Apply Lemma 2.2 for the normalized adjacency matrix M_n of $G(n,p)$ with $K = 1/\sqrt{p}$ we obtain

Corollary 2.3. *Consider the model $G(n,p)$ with $np \to \infty$ as $n \to \infty$ and let $\delta > 0$. Then for any interval $I \subset [-2,2]$ with length at least $\left(\frac{\log(np)}{\delta^4 (np)^{1/2}}\right)^{1/5}$, we have*

$$|N_I - n\int_I \rho_{sc}(x)dx| \geq \delta n \int_I \rho_{sc}(x)dx$$

with probability at most $\exp(-cn(np)^{1/2}\log(np))$.

Remark 2.4. *If one only needs the result for the bulk case $I \subset [-2+\epsilon, 2-\epsilon]$ for an absolute constant $\epsilon > 0$ then the minimum length of I can be improved to $\left(\frac{\log(np)}{\delta^4 (np)^{1/2}}\right)^{1/4}$.*

By Corollary 2.3 and Lemma 2.1, the probability that N_I fails to be close to the expected value in the model $G(n,p)$ is much smaller than the probability that $G(n,p)$ is np-regular. Thus the probability that N_I fails to be close to the expected value in the model $G_{n,d}$ where $d = np$ is the ratio of the two former probabilities, which is $O(\exp(-cn\sqrt{np}\log np))$ for some small positive constant c. Thus, Theorem 1.6 is proved, depending on Lemma 2.2 which we turn to next.

2.2 Proof of Lemma 2.2

Assume $I = [a,b]$ and $a - (-2) < 2 - b$.

We will use the approach of Guionnet and Zeitouni in [18]. Consider a random Hermitian matrix W_n with independent entries w_{ij} with support in a compact region S. Let f be a real convex L-Lipschitz function and define

$$Z := \sum_{i=1}^{n} f(\lambda_i)$$

where λ_i's are the eigenvalues of $\frac{1}{\sqrt{n}}W_n$. We are going to view Z as the function of the atom variables w_{ij}. For our application we need w_{ij} to be random variables with mean zero and variance 1, whose absolute values are bounded by a common constant K.

The following concentration inequality is from [18]

Lemma 2.5. *Let W_n, f, Z be as above. Then there is a constant $c > 0$ such that for any $T > 0$*

$$\mathbf{P}(|Z - \mathbf{E}(Z)| \geq T) \leq 4\exp(-c\frac{T^2}{K^2L^2}).$$

In order to apply Lemma 2.5 for N_I and M, it is natural to consider

$$Z := N_I = \sum_{i=1}^{n} \chi_I(\lambda_i)$$

where χ_I is the indicator function of I and λ_i are the eigenvalues of $\frac{1}{\sqrt{n}}M_n$. However, this function is neither convex nor Lipschitz. As suggested in [18], one can overcome this problem by a proper approximation. Define $I_l = [a - \frac{|I|}{C}, a]$, $I_r = [b, b + \frac{|I|}{C}]$ and construct two real functions f_1, f_2 as follows(see Figure 3):

$$f_1(x) = \begin{cases} -\frac{C}{|I|}(x-a) - 1 & \text{if } x \in (-\infty, a - \frac{|I|}{C}) \\ 0 & \text{if } x \in I \cup I_l \cup I_r \\ \frac{C}{|I|}(x-b) - 1 & \text{if } x \in (b + \frac{|I|}{C}, \infty) \end{cases}$$

$$f_2(x) = \begin{cases} -\frac{C}{|I|}(x-a) - 1 & \text{if } x \in (-\infty, a) \\ -1 & \text{if } x \in I \\ \frac{C}{|I|}(x-b) - 1 & \text{if } x \in (b, \infty) \end{cases}$$

where C is a constant to be chosen later. Note that f_j's are convex and $\frac{C}{|I|}$-Lipschitz. Define

$$X_1 = \sum_{i=1}^{n} f_1(\lambda_i), \ X_2 = \sum_{i=1}^{n} f_2(\lambda_i)$$

and apply Lemma 2.5 with $T = \frac{\delta}{8} n \int_I \rho_{sc}(t)dt$ for X_1 and X_2. Thus, we have

$$\mathbf{P}(|X_j - \mathbf{E}(X_j)| \geq \frac{\delta}{8} n \int_I \rho_{sc}(t)dt) \leq 4\exp(-c\frac{\delta^2 n^2 |I|^2 (\int_I \rho_{sc}(t)dt)^2}{K^2C^2}).$$

At this point we need to estimate the value of $\int_I \rho_{sc}(t)dt$. There are two cases: if I is in the "bulk" i.e. $I \subset [-2+\epsilon, 2-\epsilon]$ for some positive absolute constant ϵ, then $\int_I \rho_{sc}(t)dt = \alpha|I|$ where α is a constant depending on ϵ. But if I is very near the edge of $[-2, 2]$ i.e. $a-(-2) < |I| = o(1)$, then $\int_I \rho_{sc}(t)dt = \alpha'|I|^{3/2}$ for some absolute constant α'. Thus in both case we have

$$\mathbf{P}(|X_j - \mathbf{E}(X_j)| \geq \frac{\delta}{8} n \int_I \rho_{sc}(t)dt) \leq 4\exp(-c_1\frac{\delta^2 n^2 |I|^5}{K^2C^2})$$

Let $X = X_1 - X_2$, then

$$\mathbf{P}(|X - \mathbf{E}(X)| \geq \frac{\delta}{4} n \int_I \rho_{sc}(t)dt) \leq O(\exp(-c_1\frac{\delta^2 n^2 |I|^5}{K^2C^2})).$$

Now we compare X to Z, making use of a result of Götze and Tikhomirov [17]. We have $\mathbf{E}(X - Z) \leq \mathbf{E}(N_{I_l} + N_{I_r})$. In [17], Götze and Tikhomirov obtained a convergence rate for ESD of Hermitian random matrices whose entries have mean zero and variance one, which implies that for any $I \subset [-2, 2]$

$$|\mathbf{E}(N_I) - n\int_I \rho_{sc}(t)dt| < \beta n \sqrt{\frac{M_4}{n}},$$

where β is an absolute constant, $M_4 = \sup_{i,j} \mathbf{E}(|\omega_{ij}|^4)$. Thus

$$\mathbf{E}(X) \leq \mathbf{E}(Z) + n\int_{I_l \cup I_r} \rho_{sc}(t)dt + \beta n\sqrt{\frac{M_4}{n}}.$$

In the "edge" case we can choose $C = (4/\delta)^{2/3}$, then because $|I| \geq \Omega(\delta^{-2/3}(M_4/n)^{1/3})$, we have

$$n\int_{I_l \cup I_r} \rho_{sc}(t)dt = \Theta(n(\frac{|I|}{C})^{3/2}) > \Omega(n\sqrt{\frac{M_4}{n}})$$

and

$$n\int_{I_l \cup I_r} \rho_{sc}(t)dt + \beta n\sqrt{\frac{M_4}{n}} = \Theta(n(\frac{|I|}{C})^{3/2}) = \Theta(\frac{\delta}{4}n\int_I \rho_{sc}(t)dt).$$

In the "bulk" case we choose $C = 4/\delta$, then

$$n\int_{I_l \cup I_r} \rho_{sc}(t)dt + \beta n\sqrt{\frac{M_4}{n}} = \Theta(n\frac{|I|}{C}) = \Theta(\frac{\delta}{4}n\int_I \rho_{sc}(t)dt).$$

Therefore in both cases, with probability at least $1 - O(\exp(-c_1\frac{\delta^4 n^2 |I|^5}{K^2}))$, we have

$$Z \leq X \leq \mathbf{E}(X) + \frac{\delta}{4}n\int_I \rho_{sc}(t)dt < \mathbf{E}(Z) + \frac{\delta}{2}n\int_I \rho_{sc}(t)dt.$$

The convergence rate result of Götze and Tikhomirov again gives

$$\mathbf{E}(N_I) < n\int_I \rho_{sc}(t)dt + \beta n\sqrt{\frac{M_4}{n}} < (1 + \frac{\delta}{2})n\int_I \rho_{sc}(t)dt,$$

hence with probability at least $1 - O(\exp(-c_1\frac{\delta^4 n^2 |I|^5}{K^2}))$

$$Z < (1 + \delta)n\int_I \rho_{sc}(t)dt,$$

which is the desires upper bound.

The lower bound is proved using a similar argument. Let $I' = [a + \frac{|I|}{C}, b - \frac{|I|}{C}]$, $I'_l = [a, a + \frac{|I|}{C}]$, $I'_r = [b - \frac{|I|}{C}, b]$ where C is to be chosen later and define two functions g_1, g_2 as follows (see Figure 3):

$$g_1(x) = \begin{cases} -\frac{C}{|I|}(x-a) & \text{if } x \in (-\infty, a) \\ 0 & \text{if } x \in I' \cup I'_l \cup I'_r \\ \frac{C}{|I|}(x-b) & \text{if } x \in (b, \infty) \end{cases}$$

$$g_2(x) = \begin{cases} -\frac{C}{|I|}(x-a) & \text{if } x \in (-\infty, a + \frac{|I|}{C}) \\ -1 & \text{if } x \in I' \\ \frac{C}{|I|}(x-b) & \text{if } x \in (b - \frac{|I|}{C}, \infty) \end{cases}$$

Define

$$Y_1 = \sum_{i=1} g_1(\lambda_i),\ Y_2 = \sum_{i=1} g_2(\lambda_i).$$

Applying Lemma 2.5 with $T = \frac{\delta}{8} n \int_I \rho_{sc}(t)dt$ for Y_j and using the estimation for $\int_I \rho(t)dt$ as above, we have

$$\mathbf{P}(|Y_j - \mathbf{E}(Y_j)| \geq \frac{\delta}{8} n \int_I \rho_{sc}(t)dt) \leq 4\exp(-c_2 \frac{\delta^2 n^2 |I|^5}{K^2 C^2}).$$

Let $Y = Y_1 - Y_2$, then

$$\mathbf{P}(|Y - \mathbf{E}(Y)| \geq \frac{\delta}{4} n \int_I \rho_{sc}(t)dt) \leq O(\exp(-c_2 \frac{\delta^2 n^2 |I|^5}{K^2 C^2})).$$

We have $\mathbf{E}(Z - Y) \leq \mathbf{E}(N_{I'_l} + N_{I'_r})$. A similar argument as in the proof of the upper bound (using the convergence rate of Götze and Tikhomirov) shows

$$\mathbf{E}(Y) \geq \mathbf{E}(Z) - n \int_{I'_l \cup I'_r} \rho_{sc}(t)dt - \beta n \sqrt{\frac{M_4}{n}} > E(Z) - \frac{\delta}{4} n \int_I \rho_{sc}(t)dt.$$

Therefore with probability at least $1 - O(\exp(-c_2 \frac{\delta^2 n^2 |I|^5}{K^2 C^2}))$, we have

$$Z \geq Y \geq \mathbf{E}(Y) - \frac{\delta}{4} n \int_I \rho_{sc}(t)dt > \mathbf{E}(Z) - \frac{\delta}{2} n \int_I \rho_{sc}(t)dt,$$

and by the convergence rate, with probability at least $1 - O(\exp(-c2 \frac{\delta^2 n^2 |I|^5}{K^2 C^2}))$

$$Z > (1 - \delta) n \int_I \rho_{sc}(t)dt.$$

Thus, Theorem 2.2 is proved.

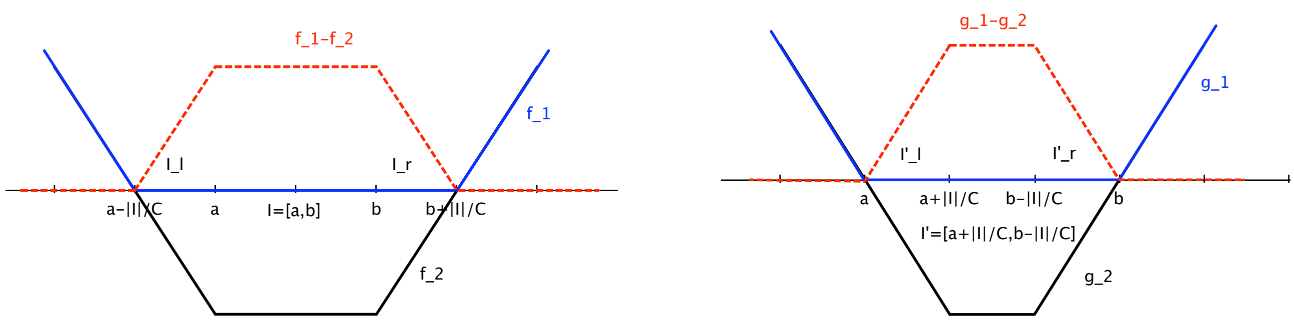

Figure 3: Auxiliary functions used in the proof

3 Infinity norm of the eigenvectors

3.1 Small perturbation lemma

A_n is the adjacency matrix of $G(n,p)$. In the proofs of Theorem 1.16 and Theorem 1.17, we actually work with the eigenvectors of a perturbed matrix

$$A_n + \epsilon N_n,$$

where $\epsilon = \epsilon(n) > 0$ can be arbitrarily small and N_n is a symmetric random matrix whose upper triangular elements are independent with a standard Gaussian distribution.

The entries of $A_n + \epsilon N_n$ are continuous and thus with probability 1, the eigenvalues of $A_n + \epsilon N_n$ are simple. Let

$$\mu_1 < \ldots < \mu_n$$

be the ordered eigenvalues of $A_n + \epsilon N_n$, which have a unique orthonormal system of eigenvectors $\{w_1, \ldots, w_n\}$. By the Cauchy interlacing principle, the eigenvalues of $A_n + \epsilon N_n$ are different from those of its principle minors, which satisfies a condition of Lemma 3.2.

Let λ_i's be the eigenvalue of A_n with multiplicity k_i defined as follows:

$$\ldots \lambda_{i-1} < \lambda_i = \lambda_{i+1} = \ldots = \lambda_{i+k_i} < \lambda_{i+k_i+1} \ldots$$

By Weyl's theorem, one has for every $1 \leq j \leq n$,

$$|\lambda_j - \mu_j| \leq \epsilon ||N_n||_{\text{op}} = O(\epsilon\sqrt{n}) \tag{3.1}$$

Thus the behaviors of eigenvalues of A_n and $A_n + \epsilon N_n$ are essentially the same by choosing ϵ sufficiently small. And everything (except Lemma 3.2) we used in the proofs of Theorem 1.16 and Theorem 1.17 for A_n also applies for $A_n + \epsilon N_n$ by a continuity argument. We will not distinguish A_n from $A_n + \epsilon N_n$ in the proofs.

The following lemma will allow us to transfer the eigenvector delocaliztion results of $A_n + \epsilon N_n$ to those of A_n at some expense.

Lemma 3.1. *In the notations of above, there exists an orthonormal basis of eigenvectors of A_n, denoted by $\{u_1, \dots, u_n\}$, such that for every $1 \le j \le n$,*

$$||u_j||_\infty \le ||w_j||_\infty + \alpha(n),$$

where $\alpha(n)$ can be arbitrarily small provided $\epsilon(n)$ is small enough.

Proof. First, since the coefficients of the characteristic polynomial of A_n are integers, there exists a positive function $l(n)$ such that either $|\lambda_s - \lambda_t| = 0$ or $|\lambda_s - \lambda_t| \ge l(n)$ for any $1 \le s, t \le n$.

By (3.1) and choosing ϵ sufficiently small, one can get

$$|\mu_i - \lambda_{i-1}| > l(n) \quad \text{and} \quad |\mu_{i+k_i} - \lambda_{i+k_i+1}| > l(n)$$

For a fixed index i, let E be the eigenspace corresponding to the eigenvalue λ_i and F be the subspace spanned by $\{w_i, \dots, w_{i+k_i}\}$. Both of E and F have dimension k_i. Let P_E and P_F be the orthogonal projection matrices onto E and F separately.

Applying the well-known Davis-Kahan theorem (see [28] Section IV, Theorem 3.6) to A_n and $A_n + \epsilon N_n$, one gets

$$||P_E - P_F||_{\text{op}} \le \frac{\epsilon ||N_n||_{\text{op}}}{l(n)} := \alpha(n),$$

where $\alpha(n)$ can be arbitrarily small depending on ϵ.

Define $v_j = P_F w_j \in E$ for $i \le j \le i + k_i$, then we have $||v_j - w_j||_2 \le \alpha(n)$. It is clear that $\{v_i, \dots, v_{k_i}\}$ are eigenvectors of A_n and

$$||v_j||_\infty \le ||w_j||_\infty + ||v_j - w_j||_2 \le ||w_j||_\infty + \alpha(n).$$

By choosing ϵ small enough such that $n\alpha(n) < 1/2$, $\{v_i, \dots, v_{k_i}\}$ are linearly independent. Indeed, if $\sum_{j=i}^{k_i} c_j v_j = 0$, one has for every $i \le s \le i + k_i$, $\sum_{j=i}^{k_i} c_j \langle P_F w_j, w_s \rangle = 0$, which implies $c_s = -\sum_{j=i}^{k_i} c_j \langle P_F w_j - w_j, w_s \rangle$. Thus $|c_s| \le \alpha(n) \sum_{j=i}^{k_i} |c_j|$, summing over all s, we can get $\sum_{j=i}^{k_i} |c_j| \le k\alpha(n) \sum_{j=i}^{k_i} |c_j|$ and therefore $c_j = 0$.

Furthermore the set $\{v_i, \ldots, v_{k_i}\}$ is 'almost' an orthonormal basis of E in the sense that

$$| \; ||v_s||_2 - 1 \; | \leq ||v_s - w_s||_2 \leq \alpha(n) \qquad \text{for any } i \leq s \leq i + k_i$$

$$\begin{aligned} |\langle v_s, v_t \rangle| &= |\langle P_F w_s, P_F w_t \rangle| \\ &= |\langle P_F w_s - w_s, P_F w_t \rangle + \langle w_s, P_F w_t - w_t \rangle| \\ &= O(\alpha(n)) \qquad \text{for any } i \leq s \neq t \leq i + k_i \end{aligned}$$

We can perform a Gram-Schmidt process on $\{v_i, \ldots, v_{k_i}\}$ to get an orthonormal system of eigenvectors $\{u_i, \ldots, u_{k_i}\}$ on E such that

$$||u_j||_\infty \leq ||w_j||_\infty + \alpha(n),$$

for every $i \leq j \leq i + k_i$.

We iterate the above argument for every distinct eigenvalue of A_n to obtain an orthonormal basis of eigenvectors of A_n.

□

3.2 Auxiliary lemmas

Lemma 3.2. *(Lemma 41, [30]) Let*

$$B_n = \begin{pmatrix} a & X^* \\ X & B_{n-1} \end{pmatrix}$$

be a $n \times n$ symmetric matrix for some $a \in \mathbb{C}$ and $X \in \mathbb{C}^{n-1}$, and let $\begin{pmatrix} x \\ v \end{pmatrix}$ be a eigenvector of B_n with eigenvalue $\lambda_i(B_n)$, where $x \in \mathbb{C}$ and $v \in \mathbb{C}^{n-1}$. Suppose that none of the eigenvalues of B_{n-1} are equal to $\lambda_i(B_n)$. Then

$$|x|^2 = \frac{1}{1 + \sum_{j=1}^{n-1} (\lambda_j(B_{n-1}) - \lambda_i(B_n))^{-2} |u_j(B_{n-1})^* X|^2},$$

where $u_j(B_{n-1})$ is a unit eigenvector corresponding to the eigenvalue $\lambda_j(B_{n-1})$.

The *Stieltjes transform* $s_n(z)$ of a symmetric matrix W is defined for $z \in \mathbb{C}$ by the formula

$$s_n(z) := \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\lambda_i(W) - z}.$$

It has the following alternate representation:

Lemma 3.3. *(Lemma 39, [30]) Let* $W = (\zeta_{ij})_{1\le i,j\le n}$ *be a symmetrix matrix, and let* z *be a complex number not in the spectrum of* W*. Then we have*

$$s_n(z) = \frac{1}{n}\sum_{k=1}^{n} \frac{1}{\zeta_{kk} - z - a_k^*(W_k - zI)^{-1}a_k}$$

where W_k *is the* $(n-1)\times(n-1)$ *matrix with the* k^{th} *row and column of* W *removed, and* $a_k \in \mathbb{C}^{n-1}$ *is the* k^{th} *column of* W *with the* k^{th} *entry removed.*

We begin with two lemmas that will be needed to prove the main results. The first lemma, following the paper [30] in Appendix B, uses Talagrand's inequality. Its proof is presented in the Appendix B.

Lemma 3.4. *Let* $Y = (\zeta_1, \ldots, \zeta_n) \in \mathbb{C}^n$ *be a random vector whose entries are i.i.d. copies of the random variable* $\zeta = \xi - p$ *(with mean* 0 *and variance* σ^2*). Let* H *be a subspace of dimension* d *and* π_H *the orthogonal projection onto* H*. Then*

$$\mathbf{P}(|\parallel \pi_H(Y)\parallel - \sigma\sqrt{d}| \ge t) \le 10\exp(-\frac{t^2}{4}).$$

In particular,

$$\parallel \pi_H(Y) \parallel = \sigma\sqrt{d} + O(\omega(\sqrt{\log n})) \tag{3.2}$$

with overwhelming probability.

The following concentration lemma for $G(n,p)$ will be a key input to prove Theorem 1.17. Let $B_n = \frac{1}{\sqrt{n}\sigma}A_n$

Lemma 3.5 (Concentration for ESD in the bulk)**.** *(Concentration for ESD in the bulk) Assume* $p = g(n)\log n/n$*. For any constants* $\varepsilon, \delta > 0$ *and any interval* I *in* $[-2+\varepsilon, 2-\varepsilon]$ *of width* $|I| = \Omega(\log^{2.2} g(n)\log n/np)$*, the number of eigenvalues* N_I *of* B_n *in* I *obeys the concentration estimate*

$$|N_I(B_n) - n\int_I \rho_{sc}(x)\,dx| \le \delta n|I|$$

with overwhelming probability.

The above lemma is a variant of Corollary 2.3. This lemma allows us to control the ESD on a smaller interval and the proof, relying on a projection lemma (Lemma 3.4), is a different approach. The proof is presented in Appendix C.

3.3 Proof of Theorem 1.16:

Let $\lambda_n(A_n)$ be the largest eigenvalue of A_n and $u = (u_1, \ldots, u_n)$ be the corresponding unit eigenvector. We have the lower bound $\lambda_n(A_n) \geq np$. And if $np = \omega(\log n)$, then the maximum degree $\Delta = (1 + o(1))np$ almost surely (See Corollary 3.14, [4]).

For every $1 \leq i \leq n$,

$$\lambda_n(A_n)u_i = \sum_{j \in N(i)} u_j,$$

where $N(i)$ is the neighborhood of vertex i. Thus, by Cauchy-Schwarz inequality,

$$||u||_\infty = \max_i \frac{|\sum_{j \in N(i)} u_j|}{\lambda_n(A_n)} \leq \frac{\sqrt{\Delta}}{\lambda_n(A_n)} = O(\frac{1}{\sqrt{np}}).$$

Let $B_n = \frac{1}{\sqrt{n}\sigma} A_n$. Since the eigenvalues of $W_n = \frac{1}{\sqrt{n}\sigma}(A_n - pJ_n)$ are on the interval $[-2, 2]$, by Lemma 1.1, $\{\lambda_1(B_n), \ldots, \lambda_{n-1}(B_n)\} \subset [-2, 2]$.

Recall that $np = g(n)\log n$. By Corollary 2.3, for any interval I with length at least $(\frac{\log(np)}{\delta^4 (np)^{1/2}})^{1/5}$(say $\delta = 0.5$),with overwhelming probability, if $I \subset [-2+\kappa, 2-\kappa]$ for some positive constant κ, one has $N_I(B_n) = \Theta(n \int_I \rho_{sc}(x)dx) = \Theta(n|I|)$; if I is at the edge of $[-2, 2]$, with length $o(1)$, one has $N_I(B_n) = \Theta(n \int_I \rho_{sc}(x)dx) = \Theta(n|I|^{3/2})$. Thus we can find a set $J \subset \{1, \ldots, n-1\}$ with $|J| = \Omega(n|I_0|)$ or $|J| = \Omega(n|I_0|^{3/2})$ such that $|\lambda_j(B_{n-1}) - \lambda_i(B_n)| \ll |I_0|$ for all $j \in J$, where B_{n-1} is the bottom right $(n-1) \times (n-1)$ minor of B_n. Here we take $|I_0| = (1/g(n)^{1/20})^{2/3}$. It is easy to check that $|I_0| \geq (\frac{\log(np)}{\delta^4 (np)^{1/2}})^{1/5}$.

By the formula in Lemma 3.2, the entry of the eigenvector of B_n can be expressed as

$$\begin{aligned}
|x|^2 &= \frac{1}{1 + \sum_{j=1}^{n-1} (\lambda_j(B_{n-1}) - \lambda_i(B_n))^{-2} |u_j(B_{n-1})^* \frac{1}{\sqrt{n}\sigma} X|^2} \\
&\leq \frac{1}{1 + \sum_{j \in J} (\lambda_j(B_{n-1}) - \lambda_i(B_n))^{-2} |u_j(B_{n-1})^* \frac{1}{\sqrt{n}\sigma} X|^2} \\
&\leq \frac{1}{1 + \sum_{j \in J} n^{-1}|I_0|^{-2} |u_j(B_{n-1})^* \frac{1}{\sigma} X|^2} = \frac{1}{1 + n^{-1}|I_0|^{-2} ||\pi_H(\frac{X}{\sigma})||^2} \\
&\leq \frac{1}{1 + n^{-1}|I_0|^{-2}|J|}
\end{aligned} \tag{3.3}$$

with overwhelming probability, where H is the span of all the eigenvectors associated to J with dimension $\dim(H) = \Theta(|J|)$, π_H is the orthogonal projection onto H and $X \in \mathbb{C}^{n-1}$ has entries that are iid copies of ξ. The last inequality in (3.3) follows from Lemma 3.4 (by taking $t = g(n)^{1/10}\sqrt{\log n}$) and the relations

$$||\pi_H(X)|| = ||\pi_H(Y + p\mathbf{1}_n)|| \geq ||\pi_{H_1}(Y + p\mathbf{1}_n)|| \geq ||\pi_{H_1}(Y)||.$$

Here $Y = X - p\mathbf{1}_n$ and $H_1 = H \cap H_2$, where H_2 is the space orthogonal to the all 1 vector $\mathbf{1}_n$. For the dimension of H_1, $\dim(H_1) \geq \dim(H) - 1$.

Since either $|J| = \Omega(n|I_0|)$ or $|J| = \Omega(n|I_0|^{3/2})$, we have $n^{-1}|I_0|^{-2}|J| = \Omega(|I_0|^{-1})$ or $n^{-1}|I_0|^{-2}|J| = \Omega(|I_0|^{-1/2})$. Thus $|x|^2 = O(|I_0|)$ or $|x|^2 = O(\sqrt{|I_0|})$. In both cases, since $|I_0| \to 0$, it follows that $|x| = o(1)$. $\square$

3.4 Proof of Theorem 1.17

With the formula in Lemma 3.2, it suffices to show the following lower bound

$$\sum_{j=1}^{n-1} (\lambda_j(B_{n-1}) - \lambda_i(B_n))^{-2} |u_j(B_{n-1})^* \frac{1}{\sqrt{n}\sigma} X|^2 \gg \frac{np}{\log^{2.2} g(n) \log n} \tag{3.4}$$

with overwhelming probability, where B_{n-1} is the bottom right $n-1 \times n-1$ minor of B_n and $X \in \mathbb{C}^{n-1}$ has entries that are iid copies of ξ. Recall that ξ takes values 1 with probability p and 0 with probability $1-p$, thus $\mathbb{E}\xi = p, \mathbb{V}ar\xi = p(1-p) = \sigma^2$.

By Theorem 3.5, we can find a set $J \subset \{1, \dots, n-1\}$ with $|J| \gg \frac{\log^{2.2} g(n) \log n}{p}$ such that $|\lambda_j(B_{n-1}) - \lambda_i(B_n)| = O(\log^{2.2} g(n) \log n/np)$ for all $j \in J$. Thus in (3.4), it is enough to prove

$$\sum_{j \in J} |u_j(B_{n-1})^T \frac{1}{\sigma} X|^2 = ||\pi_H(\frac{X}{\sigma})||^2 \gg |J|$$

or equivalently

$$||\pi_H(X)||^2 \gg \sigma^2 |J| \tag{3.5}$$

with overwhelming probability, where H is the span of all the eigenvectors associated to J with dimension $\dim(H) = \Theta(|J|)$.

Let $H_1 = H \cap H_2$, where H_2 is the space orthogonal to $\mathbf{1}_n$. The dimension of H_1 is at least $\dim(H) - 1$. Denote $Y = X - p\mathbf{1}_n$. Then the entries of Y are iid copies of ζ. By Lemma 3.4,

$$||\pi_{H_1}(Y)||^2 \gg \sigma^2 |J|$$

with overwhelming probability.

Hence, our claim follows from the relations

$$||\pi_H(X)|| = ||\pi_H(Y + p\mathbf{1}_n)|| \geq ||\pi_{H_1}(Y + p\mathbf{1}_n)|| = ||\pi_{H_1}(Y)||.$$

$\square$

Appendices

In this appendix, we complete the proofs of Theorem 1.3, Lemma 3.4 and Lemma 3.5.

A Proof of Theorem 1.3

We will show that the semicircle law holds for M_n. With Lemma 1.1, it is clear that Theorem 1.3 follows Lemma A.1 directly. The claim actually follows as a special case discussed in the paper [6]. Our proof here uses a standard moment method.

Lemma A.1. *For $p = \omega(\frac{1}{n})$, the empirical spectral distribution (ESD) of the matrix $W_n = \frac{1}{\sqrt{n}} M_n$ converges in distribution to the semicircle law which has a density $\rho_{sc}(x)$ with support on $[-2, 2]$,*

$$\rho_{sc}(x) := \frac{1}{2\pi}\sqrt{4 - x^2}.$$

Let η_{ij} be the entries of $M_n = \sigma^{-1}(A_n - pJ_n)$. For $i = j$, $\eta_{ij} = -p/\sigma$; and for $i \neq j$, η_{ij} are iid copies of random variable η, which takes value $(1-p)/\sigma$ with probability p and takes value $-p/\sigma$ with probability $1-p$.

$$\mathbf{E}\eta = 0, \mathbf{E}\eta^2 = 1, \mathbf{E}\eta^s = O\left(\frac{1}{(\sqrt{p})^{s-2}}\right) \text{ for } s \geq 2.$$

For a positive integer k, the k^{th} moment of ESD of the matrix W_n is

$$\int x^k dF_n^W(x) = \frac{1}{n}\mathbf{E}(\text{Trace}({W_n}^k)),$$

and the k^{th} moment of the semicircle distribution is

$$\int_{-2}^{2} x^k \rho_{\text{sc}}(x)dx.$$

On a compact set, convergence in distribution is the same as convergence of moments. To prove the theorem, we need to show, for every fixed number k,

$$\frac{1}{n}\mathbf{E}(\text{Trace}({W_n}^k)) \to \int_{-2}^{2} x^k \rho_{\text{sc}}(x)dx, \text{ as } n \to \infty. \tag{A.1}$$

For $k = 2m+1$, by symmetry, $\displaystyle\int_{-2}^{2} x^k \rho_{\text{sc}}(x)dx = 0$.

For $k = 2m$,

$$\begin{aligned}\int_{-2}^{2} x^k \rho_{\text{sc}}(x)dx &= \frac{1}{\pi}\int_{0}^{2} x^k \sqrt{4-x^2}dx = \frac{2^{k+2}}{\pi}\int_{0}^{\pi/2} \sin^k\theta \cos^2\theta dx \\ &= \frac{2^{k+2}}{\pi}\frac{\Gamma(\frac{k+1}{2})\Gamma(\frac{3}{2})}{\Gamma(\frac{k+4}{2})} = \frac{1}{m+1}\binom{2m}{m}\end{aligned}$$

Thus our claim (A.1) follows by showing that

$$\frac{1}{n}\mathbf{E}(\text{Trace}({W_n}^k)) = \begin{cases} O(\frac{1}{\sqrt{np}}) & \text{if } k = 2m+1; \\ \frac{1}{m+1}\binom{2m}{m} + O(\frac{1}{np}) & \text{if } k = 2m. \end{cases} \tag{A.2}$$

We have the expansion for the trace of ${W_n}^k$,

$$\begin{aligned}\frac{1}{n}\mathbf{E}(\text{Trace}({W_n}^k)) &= \frac{1}{n^{1+k/2}}\mathbf{E}(\text{Trace}(\sigma^{-1}M_n)^k) \\ &= \frac{1}{n^{1+k/2}}\sum_{1\le i_1,\dots,i_k\le n} \mathbf{E}\eta_{i_1i_2}\eta_{i_2i_3}\cdots\eta_{i_ki_1}\end{aligned} \tag{A.3}$$

Each term in the above sum corresponds to a closed walk of length k on the complete graph K_n on $\{1, 2, \dots, n\}$. On the other hand, η_{ij} are independent with mean 0. Thus the term is nonzero if and only if every edge in this closed walk appears at least twice. And we call such a walk a *good* walk. Consider a *good* walk that uses l different edges $e_1, \dots, e_l$ with corresponding multiplicities $m_1, \ldots, m_l$, where $l \leq m$, each $m_h \geq 2$ and $m_1 + \ldots + m_l = k$. Now the corresponding term to this *good* walk has form

$$\mathbf{E}\eta_{e_1}^{m_1} \cdots \eta_{e_l}^{m_l}.$$

Since such a walk uses at most $l+1$ vertices, a naive upper bound for the number of *good* walks of this type is $n^{l+1} \times l^k$.

When $k = 2m+1$, recall $\mathbf{E}\eta^s = \Theta\left((\sqrt{p})^{2-s}\right)$ for $s \geq 2$, and so

$$\begin{aligned}
\frac{1}{n}\mathbf{E}(\mathrm{Trace}({W_n}^k)) &= \frac{1}{n^{1+k/2}} \sum_{l=1}^{m} \sum_{\textit{good} \text{ walk of l edges}} \mathbf{E}\eta_{e_1}^{m_1} \cdots \eta_{e_l}^{m_l} \\
&\leq \frac{1}{n^{m+3/2}} \sum_{l=1}^{m} n^{l+1} l^k (\frac{1}{\sqrt{p}})^{m_1-2} \ldots (\frac{1}{\sqrt{p}})^{m_l-2} \\
&= O(\frac{1}{\sqrt{np}}).
\end{aligned}$$

When $k = 2m$, we classify the *good* walks into two types. The first kind uses $l \leq m-1$ different edges. The contribution of these terms will be

$$\begin{aligned}
\frac{1}{n^{1+k/2}} \sum_{l=1}^{m-1} \sum_{\text{1st kind of } \textit{good} \text{ walk of l edges}} \mathbf{E}\eta_{e_1}^{m_1} \cdots \eta_{e_l}^{m_l} &\leq \frac{1}{n^{1+m}} \sum_{l=1}^{m} n^{l+1} l^k (\frac{1}{\sqrt{p}})^{m_1-2} \ldots (\frac{1}{\sqrt{p}})^{m_l-2} \\
&= O(\frac{1}{np}).
\end{aligned}$$

The second kind of *good* walk uses exactly $l = m$ different edges and thus $m+1$ different vertices. And the corresponding term for each walk has form

$$\mathbf{E}\eta_{e_1}^2 \cdots \eta_{e_l}^2 = 1.$$

The number of this kind of *good* walk is given by the following result in the paper ([1], Page 617–618):

Lemma A.2. *The number of the second kind of good walk is*

$$\frac{n^{m+1}(1+O(n^{-1}))}{m+1}\binom{2m}{m}.$$

Then the second conclusion of (A.1) follows.

B Proof of Lemma 3.4:

The coordinates of Y are bounded in magnitude by 1. Apply Talagrand's inequality to the map $Y \to ||\pi_H(Y)||$, which is convex and 1-Lipschitz. We can conclude

$$\mathbf{P}(| \parallel \pi_H(Y) \parallel - M(\parallel \pi_H(Y) \parallel)| \geq t) \leq 4 \exp(-\frac{t^2}{16}) \tag{B.1}$$

where $M(\parallel \pi_H(Y) \parallel)$ is the median of $\parallel \pi_H(Y) \parallel$.

Let $P = (p_{ij})_{1 \leq i,j \leq n}$ be the orthogonal projection matrix onto H. One has traceP^2 =trace$P = \sum_i p_{ii} = d$ and $|p_{ii}| \leq 1$, as well as,

$$\parallel \pi_H(Y) \parallel^2 = \sum_{1 \leq i,j \leq n} p_{ij} \zeta_i \zeta_j = \sum_{i=1}^n p_{ii} \zeta_i^2 + \sum_{i \neq j} p_{ij} \zeta_i \zeta_j$$

and

$$\mathbf{E} \parallel \pi_H(Y) \parallel^2 = \mathbf{E}(\sum_{i=1}^n p_{ii} \zeta_i^2) + \mathbf{E}(\sum_{i \neq j} p_{ij} \zeta_i \zeta_j) = \sigma^2 d.$$

Take $L = 4/\sigma$. To complete the proof, it suffices to show

$$|M(\parallel \pi_H(Y) \parallel) - \sigma\sqrt{d}| \leq L\sigma. \tag{B.2}$$

Consider the event $\mathcal{E}_+$ that $\parallel \pi_H(Y) \parallel \geq \sigma L + \sigma\sqrt{d}$, which implies that $\parallel \pi_H(Y) \parallel^2 \geq \sigma^2(L^2 + 2L\sqrt{d} + d^2)$.

Let $S_1 = \sum_{i=1}^n p_{ii}(\zeta_i^2 - \sigma^2)$ and $S_2 = \sum_{i \neq j} p_{ij} \zeta_i \zeta_j$.

Now we have

$$\mathbf{P}(\mathcal{E}_+) \leq \mathbf{P}(\sum_{i=1}^n p_{ii} \zeta_i^2 \geq \sigma^2 d + L\sqrt{d}\sigma^2) + \mathbf{P}(\sum_{i \neq j} p_{ij} \zeta_i \zeta_j \geq \sigma^2 L\sqrt{d}).$$

By Chebyshev's inequality,

$$\mathbf{P}(\sum_{i=1}^{n} p_{ii}\zeta_i^2 \geq \sigma^2 d + L\sqrt{d}\sigma^2) = \mathbf{P}(S_1 \geq L\sqrt{d}\sigma^2)) \leq \frac{\mathbf{E}(|S_1|^2)}{L^2 d\sigma^4},$$

where $\mathbf{E}(|S_1|^2) = \mathbf{E}(\sum_i p_{ii}(\zeta_i^2 - \sigma^2))^2 = \sum_i p_{ii}^2 \mathbf{E}(\zeta_i^4 - \sigma^4) \leq d\sigma^2(1-2\sigma^2)$.

Therefore, $\mathbf{P}(S_1 \geq L\sqrt{d}\sigma^4) \leq \frac{d\sigma^2(1-2\sigma^2)}{L^2 d\sigma^4} < \frac{1}{16}$.

On the other hand, we have $\mathbf{E}(|S_2|^2) = \mathbf{E}(\sum_{i\neq j} p_{ij}^2 \zeta_i^2 \zeta_j^2) \leq \sigma^4 d$ and

$$\mathbf{P}(\sum_{i\neq j} p_{ij}\zeta_i\zeta_j \geq \sigma^2 L\sqrt{d}) = \mathbf{P}(S_2 \geq L\sqrt{d}\sigma^2) \leq \frac{\mathbf{E}(|S_2|^2)}{L^2 d\sigma^4} < \frac{1}{10}$$

It follows that $\mathbf{E}(\mathcal{E}_+) < 1/4$ and hence $M(\| \pi_H(Y) \|) \leq L\sigma + \sqrt{d}\sigma$.

For the lower bound, consider the event $\mathcal{E}_-$ that $\| \pi_H(Y) \| \leq \sqrt{d}\sigma - L\sigma$ and notice that

$$\mathbf{P}(\mathcal{E}_-) \leq \mathbf{P}(S_1 \leq -L\sqrt{d}\sigma^2) + \mathbf{P}(S_2 \leq -L\sqrt{d}\sigma^2).$$

The same argument applies to get $M(\| \pi_H(Y) \|) \geq \sqrt{d}\sigma - L\sigma$. Now the relations (B.1) and (B.2) together imply (3.2).

C Proof of Lemma 3.5:

Recall the normalized adjacency matrix

$$M_n = \frac{1}{\sigma}(A_n - pJ_n),$$

where $J_n = \mathbf{1}_n \mathbf{1}_n^T$ is the $n \times n$ matrix of all 1's, and let $W_n = \frac{1}{\sqrt{n}} M_n$.

Lemma C.1. *For all intervals $I \subset \mathbb{R}$ with $|I| = \omega(\log n)/np$, one has*

$$N_I(W_n) = O(n|I|)$$

with overwhelming probability.

The proof of Lemma C.1 uses the same proof as in the paper [30] with the relation (3.2).

Actually we will prove the following concentration theorem for M_n. By Lemma 1.1, $|N_I(W_n) - N_I(B_n)| \leq 1$, therefore Lemma C.2 implies Lemma 3.5.

Lemma C.2. *(Concentration for ESD in the bulk) Assume $p = g(n)\log n/n$. For any constants $\varepsilon, \delta > 0$ and any interval I in $[-2+\varepsilon, 2-\varepsilon]$ of width $|I| = \Omega(g(n)^{0.6}\log n/np)$, the number of eigenvalues N_I of $W_n = \frac{1}{\sqrt{n}}M_n$ in I obeys the concentration estimate*

$$|N_I(W_n) - n\int_I \rho_{sc}(x)\,dx| \leq \delta n|I|$$

with overwhelming probability.

To prove Theorem C.2, following the proof in [30], we consider the *Stieltjes transform*

$$s_n(z) := \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda_i(W_n) - z},$$

whose imaginary part

$$\mathrm{Im} s_n(x + \sqrt{-1}\eta) = \frac{1}{n}\sum_{i=1}^{n}\frac{\eta}{\eta^2 + (\lambda_i(W_n) - x)^2} > 0$$

in the upper half-plane $\eta > 0$.

The semicircle counterpart

$$s(z) := \int_{-2}^{2}\frac{1}{x-z}\rho_{sc}(x)\,dx = \frac{1}{2\pi}\int_{-2}^{2}\frac{1}{x-z}\sqrt{4-x^2}\,dx,$$

is the unique solution to the equation

$$s(z) + \frac{1}{s(z)+z} = 0$$

with $\mathrm{Im}s(z) > 0$.

The next proposition gives control of ESD through control of Stieltjes transform (we will take $L = 2$ in the proof):

Proposition C.3. *(Lemma 60, [30]) Let $L, \varepsilon, \delta > 0$. Suppose that one has the bound*

$$|s_n(z) - s(z)| \leq \delta$$

with (uniformly) overwhelming probability for all z with $|Re(z)| \leq L$ and $Im(z) \geq \eta$. Then for any interval I in $[-L+\varepsilon, L-\varepsilon]$ with $|I| \geq max(2\eta, \frac{\eta}{\delta}\log\frac{1}{\delta})$, one has

$$|N_I - n\int_I \rho_{sc}(x)\,dx| \leq \delta n|I|$$

with overwhelming probability.

By Proposition C.3, our objective is to show

$$|s_n(z) - s(z)| \leq \delta \tag{C.1}$$

with (uniformly) overwhelming probability for all z with $|\mathrm{Re}(z)| \leq 2$ and $\mathrm{Im}(z) \geq \eta$, where

$$\eta = \frac{\log^2 g(n)\log n}{np}.$$

In Lemma 3.3, we write

$$s_n(z) = \frac{1}{n}\sum_{k=1}^{n} \frac{1}{-\frac{\zeta_{kk}}{\sqrt{n}\sigma} - z - Y_k} \tag{C.2}$$

where

$$Y_k = a_k^*(W_{n,k} - zI)^{-1}a_k,$$

$W_{n,k}$ is the matrix W_n with the k^{th} row and column removed, and a_k is the k^{th} row of W_n with the k^{th} element removed.

The entries of a_k are independent of each other and of $W_{n,k}$, and have mean zero and variance $1/n$. By linearity of expectation we have

$$\mathbf{E}(Y_k|W_{n,k}) = \frac{1}{n}\mathrm{Trace}(W_{n,k} - zI)^{-1} = (1 - \frac{1}{n})s_{n,k}(z)$$

where

$$s_{n,k}(z) = \frac{1}{n-1}\sum_{i=1}^{n-1}\frac{1}{\lambda_i(W_{n,k}) - z}$$

is the *Stieltjes transform* of $W_{n,k}$. From the Cauchy interlacing law, we get

$$|s_n(z) - (1 - \frac{1}{n})s_{n,k}(z)| = O(\frac{1}{n}\int_{\mathbb{R}} \frac{1}{|x-z|^2}\,dx) = O(\frac{1}{n\eta}) = o(1),$$

and thus

$$\mathbf{E}(Y_k|W_{n,k}) = s_n(z) + o(1).$$

In fact a similar estimate holds for Y_k itself:

Proposition C.4. *For $1 \le k \le n$, $Y_k = \mathbf{E}(Y_k|W_{n,k}) + o(1)$ holds with (uniformly) overwhelming probability for all z with $|Re(z)| \le 2$ and $Im(z) \ge \eta$.*

Assume this proposition for the moment. By hypothesis, $|\frac{\zeta_{kk}}{\sqrt{n}\sigma}| = |\frac{-p}{\sqrt{n}\sigma}| = o(1)$. Thus in (C.2), we actually get

$$s_n(z) + \frac{1}{n}\sum_{k=1}^{n} \frac{1}{s_n(z) + z + o(1)} = 0 \tag{C.3}$$

with overwhelming probability. This implies that with overwhelming probability either $s_n(z) = s(z) + o(1)$ or that $s_n(z) = -z + o(1)$. On the other hand, as $\mathrm{Im} s_n(z)$ is necessarily positive, the second possibility can only occur when $\mathrm{Im} z = o(1)$. A continuity argument (as in [11]) then shows that the second possibility cannot occur at all and the claim follows.

Now it remains to prove Proposition C.4.

Proof of Proposition C.4. Decompose

$$Y_k = \sum_{j=1}^{n-1} \frac{|u_j(W_{n,k})^* a_k|^2}{\lambda_j(W_{n,k}) - z}$$

and evaluate

$$\begin{aligned} Y_k - \mathbf{E}(Y_k|W_{n,k}) &= Y_k - (1 - \frac{1}{n})s_{n,k}(z) + o(1) \\ &= \sum_{j=1}^{n-1} \frac{|u_j(W_{n,k})^* a_k|^2 - \frac{1}{n}}{\lambda_j(W_{n,k}) - z} + o(1) \\ &= \sum_{j=1}^{n-1} \frac{R_j}{\lambda_j(W_{n,k}) - z} + o(1), \end{aligned} \tag{C.4}$$

where we denote $R_j = |u_j(W_{n,k})^* a_k|^2 - \frac{1}{n}$, $\{u_j(W_{n,k})\}$ are orthonormal eigenvectors of $W_{n,k}$.

Let $J \subset \{1, \dots, n-1\}$, then

$$\sum_{j \in J} R_j = ||P_H(a_k)||^2 - \frac{\dim(H)}{n}$$

where H is the space spanned by $\{u_j(W_{n,k})\}$ for $j \in J$ and P_H is the orthogonal projection onto H.

In Lemma 3.4, by taking $t = h(n)\sqrt{\log n}$, where $h(n) = \log^{0.001} g(n)$, one can conclude with overwhelming probability

$$|\sum_{j \in J} R_j| \ll \frac{1}{n} \left(\frac{h(n)\sqrt{|J| \log n}}{\sqrt{p}} + \frac{h(n)^2 \log n}{p} \right). \tag{C.5}$$

Using the triangle inequality,

$$\sum_{j \in J} |R_j| \ll \frac{1}{n} \left(|J| + \frac{h(n)^2 \log n}{p} \right) \tag{C.6}$$

with overwhelming probability.

Let $z = x + \sqrt{-1}\eta$, where $\eta = \log^2 g(n) \log n / np$ and $|x| \leq 2 - \varepsilon$, define two parameters

$$\alpha = \frac{1}{\log^{4/3} g(n)} \qquad \text{and} \qquad \beta = \frac{1}{\log^{1/3} g(n)}.$$

First, for those $j \in J$ such that $|\lambda_j(W_{n,k}) - x| \leq \beta\eta$, the function $\frac{1}{\lambda_j(W_{n,k}) - x - \sqrt{-1}\eta}$ has magnitude $O(\frac{1}{\eta})$. From Lemma C.1, $|J| \ll n\beta\eta$, and so the contribution for these $j \in J$ is,

$$|\sum_{j \in J} \frac{R_j}{\lambda_j(W_{n,k}) - z}| \ll \frac{1}{n\eta} \left(n\beta\eta + \frac{h(n)^2}{\log^2 g(n)} \right) = O(\frac{1}{\log^{1/3} g(n)}) = o(1).$$

For the contribution of the remaining $j \in J$, we subdivide the indices as

$$a \le |\lambda_j(W_{n,k}) - x| \le (1+\alpha)a$$

where $a = (1+\alpha)^l \beta\eta$, for $0 \le l \le L$, and then sum over l.

For each such interval, the function $\frac{1}{\lambda_j(W_{n,k})-x-\sqrt{-1}\eta}$ has magnitude $O(\frac{1}{a})$ and fluctuates by at most $O(\frac{\alpha}{a})$. Say J is the set of all j's in this interval, thus by Lemma C.1, $|J| = O(n\alpha a)$. Together with bounds (C.5), (C.6), the contribution for these j on such an interval,

$$\begin{aligned}
|\sum_{j\in J} \frac{R_j}{\lambda_j(W_{n,k}) - z}| &\ll \frac{1}{an}\left(\frac{h(n)\sqrt{|J|\log n}}{\sqrt{p}} + \frac{h(n)^2 \log n}{p}\right) + \frac{\alpha}{an}\left(|J| + \frac{h(n)^2\log n}{p}\right) \\
&= O\left(\frac{\sqrt{\alpha}}{\sqrt{(1+\alpha)^l}}\frac{h(n)}{\sqrt{\beta}\log g(n)} + \frac{h^2(n)}{(1+\alpha)^l \beta \log^2 g(n)} + \alpha^2\right) \\
&= O\left(\frac{1}{\sqrt{\alpha\beta}}\frac{h(n)}{\log g(n)} + \alpha \log \frac{1}{\beta\eta}\right)
\end{aligned}$$

Summing over l and noticing that $(1+\alpha)^L \eta / g(n)^{1/4} \le 3$, we get

$$\begin{aligned}
|\sum_{j\in J, \mathrm{all} J} \frac{R_j}{\lambda_j(W_{n,k}) - z}| &= O\left(\frac{1}{\sqrt{\alpha\beta}}\frac{h(n)}{\log g(n)} + \alpha\log\frac{1}{\beta\eta}\right) \\
&= O\left(\frac{h(n)}{\log^{1/6} g(n)}\right) = o(1).
\end{aligned}$$

□

Acknowledgement. The authors thank Terence Tao for useful conversations.

References

[1] Z.D. Bai. Methodologies in spectral analysis of large dimensional random matrices, a review. In *Advances in statistics: proceedings of the conference in honor of Professor Zhidong Bai on his 65th birthday, National University of Singapore, 20 July 2008*, volume 9, page 174. World Scientific Pub Co Inc, 2008.

[2] M. Bauer and O. Golinelli. Random incidence matrices: moments of the spectral density. *Journal of Statistical Physics*, 103(1):301–337, 2001.

[3] S. Bhamidi, S.N. Evans, and A. Sen. Spectra of large random trees. *Arxiv preprint arXiv:0903.3589*, 2009.

[4] B. Bollobás. *Random graphs*. Cambridge Univ Pr, 2001.

[5] S. Brooks and E. Lindenstrauss. Non-localization of eigenfunctions on large regular graphs. *Arxiv preprint arXiv:0912.3239*, 2009.

[6] F. Chung, L. Lu, and V. Vu. The spectra of random graphs with given expected degrees. *Internet Mathematics*, 1(3):257–275, 2004.

[7] Y. Dekel, J. Lee, and N. Linial. Eigenvectors of random graphs: Nodal domains. In *APPROX '07/RANDOM '07: Proceedings of the 10th International Workshop on Approximation and the 11th International Workshop on Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 436–448, Berlin, Heidelberg, 2007. Springer-Verlag.

[8] Y. Dekel, J. Lee, and N. Linial. Eigenvectors of random graphs: Nodal domains. *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 436–448, 2008.

[9] I. Dumitriu and S. Pal. Sparse regular random graphs: spectral density and eigenvectors. *Arxiv preprint arXiv:0910.5306*, 2009.

[10] L. Erdos, B. Schlein, and H.T. Yau. Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices. *Ann. Probab*, 37(3):815–852, 2009.

[11] L. Erdos, B. Schlein, and H.T. Yau. Local semicircle law and complete delocalization for Wigner random matrices. Accepted in Comm. Math. Phys. *Communications in Mathematical Physics*, 287(2):641–655, 2010.

[12] U. Feige and E. Ofek. Spectral techniques applied to sparse random graphs. *Random Structures and Algorithms*, 27(2):251, 2005.

[13] J. Friedman. On the second eigenvalue and random walks in random d-regular graphs. *Combinatorica*, 11(4):331–362, 1991.

[14] J. Friedman. Some geometric aspects of graphs and their eigenfunctions. *Duke Math. J*, 69(3):487–525, 1993.

[15] J. Friedman. A proof of Alon's second eigenvalue conjecture. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 720–724. ACM, 2003.

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

[17] F. Götze and A. Tikhomirov. Rate of convergence to the semi-circular law. *Probability Theory and Related Fields*, 127(2):228–276, 2003.

[18] A. Guionnet and O. Zeitouni. Concentration of the spectral measure for large matrices. *Electron. Comm. Probab*, 5:119–136, 2000.

[19] S. Janson, T. Łuczak, and A. Ruciński. *Random graphs.* Citeseer, 2000.

[20] M. Krivelevich, B. Sudakov, V.H. Vu, and N.C. Wormald. Random regular graphs of high degree. *Random Structures and Algorithms*, 18(4):346–363, 2001.

[21] B.D. McKay. The expected eigenvalue distribution of a large regular graph. *Linear Algebra and its Applications*, 40:203–216, 1981.

[22] B.D. McKay and N.C. Wormald. Asymptotic enumeration by degree sequence of graphs with degrees $o(n^{1/2})$. *Combinatorica 11 , no. 4, 369–382.*, 1991.

[23] B.D. McKay and N.C. Wormald. The degree sequence of a random graph. I. The models. *Random Structures Algorithms 11, no. 2, 97–117*, 1997.

[24] A. Pothen, H.D. Simon, and K.P. Liou. Partitioning sparse matrices with eigenvectors of graphs. *SIAM Journal on Matrix Analysis and Applications*, 11:430, 1990.

[25] G. Semerjian and L.F. Cugliandolo. Sparse random matrices: the eigenvalue spectrum revisited. *Journal of Physics A: Mathematical and General*, 35:4837–4851, 2002.

[26] E Shamir and E Upfal. Large regular factors in random graphs. *Convexity and graph theory (Jerusalem, 1981)*, 1981.

[27] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 22(8):888–905, 2000.

[28] G.W. Stewart and Ji-guang. Sun. *Matrix perturbation theory.* Academic press New York, 1990.

[29] T. Tao and V. Vu. Random matrices: Universality of local eigenvalue statistics up to the edge. *Communications in Mathematical Physics*, pages 1–24, 2010.

[30] T. Tao and V. Vu. Random matrices: Universality of the local eigenvalue statistics, submitted. *Institute of Mathematics, University of Munich, Theresienstr*, 39, 2010.

[31] V. Vu. Random discrete matrices. *Horizons of Combinatorics*, pages 257–280, 2008.

[32] E.P. Wigner. On the distribution of the roots of certain symmetric matrices. *Annals of Mathematics*, 67(2):325–327, 1958.

[33] N.C. Wormald. Models of random regular graphs. *London Mathematical Society Lecture Note Series*, pages 239–298, 1999.