

Eigenvectors of Random Graphs : Delocalization and Nodal Domains

Sanjeev Arora and Aditya Bhaskara

Department of Computer Science
Princeton University, Princeton NJ 08540
{arora, bhaskara}@cs.princeton.edu

Abstract. We study properties of the eigenvectors of adjacency matrices of $G(n, p)$ random graphs, for $p = \omega(\log n)/n$. This connects to similar investigations for other random matrix models studied in physics and mathematics. Motivated by the recent paper of Dekel, Lee and Linial we study *delocalization* properties of eigenvectors and their connection to *nodal domains*. We show the following for an eigenvector x (normalized s.t. $\|x\|_2 = 1$):

1. For any $S \subseteq [n]$ and $|S| = \delta n$, we have

$$\sum_{i \in S} x_i^2 \geq \frac{\delta^4 p^6}{\log^2(1/p\delta)} \quad w.h.p.$$

A similar statement proved for $\delta > 1/2$ by Dekel, Lee and Linial.

2. Let $p > n^{-1/20}$. Then x has exactly two nodal domains whp. (i.e., the subgraph of vertices with $x_i \geq 0$ is connected and so is the subgraph of vertices with $x_i \leq 0$). Previously such a statement was not known even for $p = 1/2$, unless one is allowed to discard $O(1/p)$ “exceptional” vertices of the graph.

Our techniques involve using Wigner’s semicircle law on “short scales”, an idea previously used in mathematical physics by Erdős, Schlein and Yau.

1 Introduction

Eigenvalues and eigenvectors of random matrices is a topic that spans multiple disciplines of science, and is related to topics as diverse as the zeros of the Riemann Zeta function, properties of hamiltonians of atomic nuclei, and conductance properties of metallic lattices. The seminal work of Wigner [1] studied matrices whose entries are i.i.d. (standard) gaussians, and identified the distribution of eigenvalues, which he quantified in the celebrated *semi-circle law*. Our knowledge of the finer distribution of eigenvalues has grown in the decades since, and extensions of these results to other random matrix models have been found. In this paper we are interested in adjacency matrices of random graphs in the Erdős-Renyi model $G(n, p)$ wherein each edge is picked iid with probability p . The eigenvalue results in Wigner tradition have also been proved for it (see [2]).

Establishing the properties of *eigenvectors* has proved much harder than for *eigenvalues* in most models. Hints from a variety of disciplines suggest that eigenvectors of random matrices should look like random vectors in \mathfrak{R}^n . A natural conjecture is that the entries are distributed like i.i.d. Gaussians. This is easy to prove via a symmetry argument when the random matrix has i.i.d. Gaussian entries (note that this distribution is invariant under rotation), and with some effort weaker results are provable for the case of Gaussian Orthogonal Ensembles (we refer to the excellent survey of [3]). Recently Tao and Vu [4] and Erdős, Schlein and Yau [5] have proved strong theorems about distributions of eigenvectors for more general distributions.

When a full characterization of eigenvector properties eludes us, we may try to prove *delocalization* properties of eigenvectors as a first step. Delocalization says that the “mass distribution” over the coordinates of the eigenvectors is not too lumpy. If the eigenvector entries were truly Gaussian, one would expect entries to be about $1/\sqrt{n}$ in magnitude, so the following is a small list of properties one may expect to hold with high probability (throughout the paper, “whp.” will mean with probability bigger than $1 - n^{-c}$ for a large constant c):

1. $\|x\|_4 \leq \log n$ (Posed by T. Spencer, see [3].)
2. $\|x\|_\infty \leq \frac{\log n}{\sqrt{n}}$. (Stronger than the previous statement.)
3. $|x \cdot u| \leq \frac{\log n}{\sqrt{n}}$ w.h.p. for every unit vector u .
4. Say $S \subseteq [n]$ and $|S| = \delta n$. Then $\sum_{i \in S} x_i^2$ is bounded from above and below by ‘reasonable’ functions of δ w.h.p.

Dekel, Lee and Linial [6] were led to some of the above questions because of their study of *nodal domains*. Given an $f : V \mapsto \mathfrak{R}$, its nodal domains are defined as the connected components formed by vertices having the same sign for f (some care is needed in dealing with v s.t. $f(v) = 0$, which we will get to later). Nodal domains arising out of physical systems have been studied extensively. For a review of the subject, we refer to the articles in [7]. For instance, Blum *et al* have shown that nodal count statistics can be used as a criterion for quantum chaos. Closer to home, in computer graphics they are used as a way to distinguish between surfaces whose associated graphs happen to be iso-spectral (see [8] for these and other examples).

Courant’s celebrated *nodal domain theorem* (c.f. [7]) says that the k th eigenvector (after sorting based on increasing eigenvalues) of the Laplacian of a graph has at most k nodal domains. A *canonical* example is that of harmonics on a string: the k th eigenfunction is a sine function with k peaks and k sign changes. Connections are also known between nodal domains and chromatic number (we refer to the survey [9]).

In $G(n, p)$ graphs which are *dense enough*, we do not expect the eigenvectors to have more than two nodal domains (one each for positive and negative). The intuition is that each vertex with a negative value in the eigenvector is very likely to have an edge to a vertex in the domain corresponding to negative vertices, so the entire negative domain should just be connected. Dekel, Lee and Linial [6] studied this question, and proved that this is true, provided we throw away

$O(1/p)$ “exceptional” vertices. As p gets smaller, this number is quite large, and thus they cannot obtain a good bound on the number of nodal domains.

Our results. We prove that there are precisely two nodal domains in any eigenvector of $G(n, p)$ whp., for $p \geq n^{-1/19+\varepsilon}$ (any $\varepsilon > 0$). Our proof has two key ingredients: the first is a strong delocalization result (in this case an ℓ_∞ bound) on the eigenvectors proved very recently due to [10].¹ The second is a stronger “spreading” result we prove – it says that for any subset S of δn coordinates, an eigenvector has non-trivial ℓ_2 mass on S whp. Dekel, Lee and Linial [6] also prove such a result, however they can do this only for $\delta > 1/2$.

The emphasis of our results is on the connection between *delocalization* (in this case, saying that mass is “spread-out”, even to sets of a small constant size), and properties of nodal domains. Our techniques are inspired by the work of Erdős, Schlein and Yau [5] who give a beautiful technique to prove properties of eigenvectors starting with refined versions of the semicircle law. (Section 2.1) A key component of their argument is proving that the gap between two eigenvalues is small for *most* of the spectrum – it is only around $1/n$, if eigenvalues are normalized to lie in $[-2, 2]$. Gaps between eigenvalues are interesting in their own right, and are related to, for instance, the distribution of roots of the Riemann zeta function [3]. We rely on results of this flavor for $G(n, p)$, proved recently in work by Tran *et al* [11], and Erdős *et al* [10].

Outline. The delocalization result is proved in Section 3. This is used to prove the theorem on nodal domains (Section 4). We end with a discussion on other natural delocalization properties and directions for future work.

2 Notation and Preliminaries

We consider the Erdős-Renyi graph model $G(n, p)$, in which we have n vertices and each possible edge is picked independently with probability p . We study properties of the adjacency matrix of $G \sim G(n, p)$.

Define \mathcal{D}_p to be a random variable which takes value 1 w.p. p and value 0 w.p. $1 - p$. Define \mathcal{X}_p to be the “shifted” variant: it takes value $1 - p$ w.p. p , and $-p$ w.p. $1 - p$ (and hence has a mean zero). Note that the adjacency matrix has zero on the diagonal, other entries are distributed according to \mathcal{D}_p (and they are independent, modulo symmetry constraints).

Let $\mathbf{1}_n$ denote the vector in \mathfrak{R}^n with every coordinate (in the standard basis) equal to $1/\sqrt{n}$ (thus it is a unit vector for each n). Also, $J_{m,n}$ will refer to an $m \times n$ matrix with each entry equal to 1. Finally, whenever we say that a statement is true ‘whp.’, we mean that it is true with probability $1 - \frac{1}{\text{poly}(n)}$. As is standard, there will be constants which can be chosen so as to make the statements true for any given polynomial. We also take union bounds over polynomially many events that are true whp., and the result will be true whp. Let us now outline some standard results from random matrix theory.

¹ An earlier version of this paper used a weaker bound due to [11].

2.1 Semicircle law and the eigenvalues of $G(n, p)$

Let G be drawn from $G(n, p)$, with $p > \omega(\log n)/n$ and let A denote its adjacency matrix. The following facts are well known about the eigenvalues of A (we refer to the excellent lecture notes due to Tao [12] for details):

Fact 1 *The largest eigenvalue λ_{max} is $np(1 \pm O(\frac{1}{\sqrt{n}}))$. Further, the eigenvector corresponding to the largest eigenvalue is ‘close’ to the vector $\mathbf{1}_n$ (see [6] or [13]).*

Fact 2 *The rest of the eigenvalues (apart from the first) satisfy $|\lambda_i| \leq 2(1 + o(1))\gamma\sqrt{n}$ ($= 2(1 + o(1))\sqrt{np(1-p)}$).*

Fact 3 (Semicircle law) *Let $\varepsilon > 0$ and $t, t + \varepsilon \in [-2, 2]$. Let $N(t, \varepsilon)$ denote the number of eigenvalues of A that lie between $t\gamma\sqrt{n}$ and $(t + \varepsilon)\gamma\sqrt{n}$. Then for any $\delta > 0$, we have*

$$\left| N(t, \varepsilon) - n \int_t^{t+\varepsilon} \rho_{sc}(x) dx \right| \leq \delta n \int_t^{t+\varepsilon} \rho_{sc}(x) dx$$

whp., where $\rho_{sc}(x)$ denotes the ‘semicircle density’: $\frac{1}{2\pi}\sqrt{4-x^2}$.

A lot of work has been aimed at proving the semicircle law for ε (in the statement above) as small as possible. These are called *local* semicircle laws, and play a crucial role in all the existing results on properties of eigenvectors ([5],[11],[10]).

2.2 Singular values of random matrices

The distribution of singular values of random rectangular matrices has also been extensively studied in literature. We consider rectangular $m \times n$ matrices (think of $m > n$), with entries drawn iid from a distribution \mathcal{D} with mean zero and variance γ^2 . Bai and Yin [14] showed asymptotic (in m, n , with fixed ratio m/n) bounds on the smallest and largest singular values, for fairly general \mathcal{D} . This was extended to more general m, n in later works, culminating in the result of Rudelson and Vershynin [15], which we state for the special case of $\mathcal{D} = \mathcal{X}_p$ (defined earlier in this section).

Theorem 1. [15] *Suppose A is an $m \times n$ matrix with entries drawn iid from \mathcal{X}_p , and let $\gamma^2 = p(1-p)$, as earlier. Then we have*

$$\sigma_{\max}(A) = (1+o(1))\gamma(\sqrt{m}+\sqrt{n}), \quad \text{and} \quad \sigma_{\min}(A) = (1-o(1))\gamma(\sqrt{m}-\sqrt{n}) \text{ whp.}$$

We will only need fairly innocuous applications of this theorem – for instance the $o(1)$ terms could be, say $1/2$. Such versions are in fact much easier to prove.

3 Support on δ -sized sets

In what follows, let G be a graph drawn from $\mathcal{G}(n, p)$, and let A denote its adjacency matrix. Let us fix $1 < i \leq n$, and let $\mathbf{v} \in \mathbb{R}^n$ be the eigenvector corresponding to the i th largest eigenvalue of A .² The main result of this section considers the distribution of *mass* on this eigenvector. More precisely,

Theorem 2. *Let G, A and \mathbf{v} be defined as in the paragraph above, and let $S \subseteq [n]$ be a set of indices satisfying $|S| \geq \delta n$, for some parameter $\delta > 0$. Then we have*

$$\|\mathbf{v}_{|S}\|_2^2 \geq \Omega\left(\frac{\delta^4 p^6}{\log^2(1/p\delta)}\right) \quad \text{whp.},$$

where $\|\mathbf{v}_{|S}\|_2^2 := \sum_{j \in S} \mathbf{v}_j^2$.

Comments. Note that the “whp.” is after fixing an S . It would be nice to say that *every* set of size δn has a large mass, but we do not know how to do this. Another point is that the bound degrades (quite badly) as p gets smaller – this is a common issue with $G(n, p)$ and it is not clear how it can be removed.

As a warm-up, we will consider a simpler case. Here we will look at a symmetric matrix with entries drawn i.i.d. from the distribution \mathcal{X}_p (instead of an adjacency matrix). More precisely,

Lemma 1. *Let M be an $n \times n$ symmetric matrix with the upper diagonal entries drawn i.i.d. from the (symmetric) distribution \mathcal{X}_p . Let $S \subseteq [n]$ be a set of indices, with $|S| > \delta n$, and $\delta > 1/2$. Lastly, let \mathbf{v} be an eigenvector of M , normalized to have unit length. Then w.h.p., we have $\|\mathbf{v}_{|S}\|_2^2 \geq \Omega(1)$.*

Having entries from \mathcal{X}_p and not \mathcal{D}_p simplifies things. The other (more crucial) simplification is in assuming that δ is a constant $> 1/2$. The difference will determine the constant in the $\Omega(\cdot)$. This is the key lemma in the work of Dekel *et al.*

Proof. Let λ be the eigenvalue corresponding to \mathbf{v} . Now let us view the matrix M as *blocks* based on indices being in S and $V \setminus S$. We write the equation $M\mathbf{v} = \lambda\mathbf{v}$ as

$$\begin{bmatrix} B & H \\ H^T & D \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \lambda \begin{bmatrix} x \\ y \end{bmatrix}, \quad (1)$$

(To be clear, the vector \mathbf{v} is split into vectors x, y based on indices in $S, V \setminus S$, and matrix M is split into B, H, H^T, D , since it is symmetric). Let us denote $|S| = m$ ($= \delta n$), and let $r = n - m$; thus the dimensions of B, H, D are $m \times m, m \times r, r \times r$ respectively. v is normalized, thus $\|x\|^2 + \|y\|^2 = 1$. From (1), we have

$$\begin{aligned} Bx + Hy &= \lambda x, \\ H^T x + Dy &= \lambda y. \end{aligned} \quad (2)$$

² We do not deal with $i = 1$ since it is very simple – in this case standard results (c.f. [13]) say that the eigenvector is close (entry-wise) to the all-ones vector (w.h.p.), and all our claims follow from this.

Assumption. We will assume that neither B nor D has λ as an eigenvalue, and will thus freely use the expression $(\lambda I - B)^{-1}$ and the like. This is a technical issue – we can deal with it by either adding a small amount of noise (as in Section 3.1 of [11]), or by working with the pseudo-inverse.

Now from the first equation of (2), we have that $x = (\lambda I - B)^{-1}Hy$. Since $\|x\|^2 + \|y\|^2 = 1$, and we are done if $\|x\|^2 > 1/4$, we may assume that $\|y\|^2 > 3/4$. The key now is that for $\delta > 1/2$, H is an $m \times r$ matrix, with $m > cr$ for some constant $c > 1$. Thus H is well-conditioned, more precisely (see Section 2), we have that

$$\sigma_{\min}(H) \geq \gamma(\sqrt{m} - \sqrt{r}) > \gamma\sqrt{m} \cdot c',$$

for some constant c' (depending on c). Since $\|Hz\|^2 \geq \sigma_{\min}^2 \|z\|^2$ for any z , we have in particular that $\|Hy\|^2 \geq m\gamma^2 \cdot c'^2$.

We will claim that the largest eigenvalue (in magnitude) of $(\lambda I - B)$ is at most $6\gamma\sqrt{m}$. This would imply that the *smallest* eigenvalue of $(\lambda I - B)^{-1}$ is at least $\frac{1}{6\gamma\sqrt{m}}$ in magnitude, thus for any vector z , $\|(\lambda I - B)^{-1}z\|^2 \geq \frac{1}{36\gamma^2 m} \|z\|^2$, so plugging in $z = Hy$ and using the earlier bound, the lemma follows.

It thus remains to verify that the largest eigenvalue of $\lambda I - B$ is at most $6\gamma\sqrt{m}$. First, note that λ is an eigenvalue of M , a symmetric matrix with entries i.i.d. copies of the symmetric r.v. \mathcal{X}_p . Thus $\lambda < 2\gamma\sqrt{n} < 3\gamma\sqrt{m}$ (since $m > n/2$). Second, since B is a symmetric random matrix with entries from \mathcal{X}_p , the eigenvalues of B all have magnitude smaller than $3\gamma\sqrt{n}$ w.h.p. These imply the desired bound.

General case. The main difficulty in the general case ($\delta \leq 1/2$) is that we no longer have a good lower bound on $\|Hy\|^2$ – indeed such a bound is false for arbitrary y (there will even exist non-zero y s.t. $Hy = 0$!). Another (more technical) difficulty is that we wish to deal with the adjacency matrix, whose entries are distributed according to \mathcal{D}_p and not the symmetric \mathcal{X}_p : this means that $(\lambda I - B)^{-1}$ will have one eigenvalue which is *tiny* (of the order of $1/n$ if we think of p as a constant), and we need to ensure that Hy does not have most of its mass long this eigenvector.

Outline. For the sake of intuition, suppose the second issue did not exist (i.e., we are dealing with symmetric matrices with upper-diagonal entries i.i.d. from \mathcal{X}_p). Then the first issue is dealt with by observing the following: even though $\|Hy\|$ could be small for certain y , it is still true that given y , this quantity is large with high probability (over the choice of H). So the key idea is to try to restrict the possible y to a “small” set (which does not depend on H), and then take a union bound. The restriction is achieved by observing that the second equality in (2) implies that $y = (\lambda I - D)^{-1}H^T x$, and if x is “too short”, then so is $H^T x$, and hence $(\lambda I - D)^{-1}$ should “stretch” $H^T x$ by a large amount (because y is close to length 1). This in turn implies that y is *essentially* supported on the large singular vectors of $(\lambda I - D)^{-1}$ (which is independent of the choice of H !). This allows the union bound argument to go through.

Dealing with the second issue above, namely working with \mathcal{D}_p will involve many technical difficulties, but the rough outline is as described above: first

analyze for a fixed y , then impose a “structure” on y , followed by an ε -net argument.

3.1 Theorem 2: Analyzing a fixed y

Let us now show, for a given vector y , how $(\lambda I - B)^{-1}Hy$ has a large norm w.h.p. (over choice of H). This step is more or less straightforward if the entries of the matrix B i.i.d. from \mathcal{X}_p . However in our case B is the adjacency matrix of the induced (also random with edge prob. p) graph on S . It has mp as its top eigenvalue and all other eigenvalues are $< 3\gamma\sqrt{m}$ whp. Since the parameter $\lambda < 3\gamma\sqrt{n}$, we have that all except possibly one, of the eigenvalues of $(\lambda I - B)$ have magnitude at most $6\gamma\sqrt{n}$. Let $u \in \mathfrak{R}^m$ be the eigenvector corresponding to the only “large” ($> 6\gamma\sqrt{n}$) eigenvalue. Now if the vector Hy has “enough” mass orthogonal to u , it turns out we can give a lower bound on the length of $(\lambda I - B)^{-1}Hy$. We thus show the following.

Lemma 2. *Suppose $u \in \mathfrak{R}^m$ is a unit vector s.t. $\|u - \mathbf{1}_m\|^2 < 1$, and let $y \in \mathfrak{R}^r$. Let H be a matrix in $\mathfrak{R}^{m \times r}$ whose entries are i.i.d. copies of \mathcal{D}_p (it is rectangular – we do not have any symmetry conditions). Then we have*

$$\Pr [\|Hy\|^2 - \langle u, Hy \rangle^2 < mp\gamma^2\|y\|^2/10] < e^{-mp\gamma^2/200}.$$

Proof. First, let us write $H = H' + pJ_{m,r}$, as before. Also since the statement is invariant under scaling y , let us assume $\|y\| = 1$. Further, let $y = y' + \alpha\mathbf{1}_r$, where $\langle y, \mathbf{1}_r \rangle = 0$. Then $Hy = H'y + (mr)^{1/2}p\alpha \cdot \mathbf{1}_m$. Let us write $c := (mr)^{1/2}p\alpha$, and let $Z = H'y$. Thus for random H , Z is a vector with i.i.d. mean-zero entries. The quantity we are interested in is now

$$\begin{aligned} & \|Z + c\mathbf{1}_m\|^2 - \langle u, (Z + c\mathbf{1}_m) \rangle^2 \\ &= \|Z\|^2 + \|c\mathbf{1}_m\|^2 + 2\langle Z, c\mathbf{1}_m \rangle - (\langle u, Z \rangle^2 + \langle u, c\mathbf{1}_m \rangle^2 + 2\langle u, Z \rangle \langle u, c\mathbf{1}_m \rangle) \\ &= \left(\|Z\|^2 - \langle u, Z \rangle^2 \right) + \left(\langle \mathbf{1}_m, c\mathbf{1}_m \rangle^2 - \langle u, c\mathbf{1}_m \rangle^2 + 2\langle \mathbf{1}_m, Z \rangle \langle \mathbf{1}_m, c\mathbf{1}_m \rangle - 2\langle u, Z \rangle \langle u, c\mathbf{1}_m \rangle \right) \end{aligned}$$

Let \mathcal{T}_1 and \mathcal{T}_2 denote the terms in the first and the second parentheses. The idea is now to manipulate the terms, \mathcal{T}_1 and \mathcal{T}_2 separately and using Azuma’s inequality to prove that they are both large, whp. The details are somewhat tricky, though morally straightforward. The entire proof of the lemma is reproduced in the appendix (Section A.2). \square

3.2 Theorem 2: Imposing structure on y

First, we show how to restrict the possible set of y . To this effect, we show the following:

Lemma 3. *Let $x \in \mathfrak{R}^m$, $y \in \mathfrak{R}^n$, with $\|x\| < \varepsilon^2 d/2n$, and suppose $\|y\| > 3/4$. Further, let $y = (\lambda I - D)^{-1}H^T x$ (with H, D as before). Then there exist a set S of at most d vectors, determined entirely by D , such that $y = s + z$, with $s \in \text{span}(S)$ and $\|z\| < \varepsilon$.*

Before going into the proof, let us show: (proof in Section A.3)

Lemma 4. *Suppose $M \in \mathfrak{R}^{n \times n}$ and $u, v \in \mathfrak{R}^n$, with $v = Mu$. Suppose $\|v\| = 1$, and $\|u\| < \varepsilon/\tau$, for some $\varepsilon, \tau > 0$. Then $v = v' + z$, where $\|z\| < \varepsilon$, and v' is in the span of the eigenvalues of M of magnitude $> \tau$.*

Proof ((of Lemma 3)). Let us write $H = H' + pJ_{m,r}$, so we get

$$y = (\lambda I - D)^{-1}(H'^T + pJ_{r,m})x = (\lambda I - D)^{-1}H'^T x + p(\lambda I - D)^{-1}J_{r,m}x$$

The second term is a vector which is parallel to $(\lambda I - D)^{-1}\mathbf{1}_r$. Let us add a unit vector along this direction to S . Now consider the vector $\eta := (\lambda I - D)^{-1}H'^T x$. If $\|\eta\| < \varepsilon$, we are done: we have expressed y as the span of a small set of vectors (which depend only on D) plus an ε noise. Thus suppose $\|\eta\| > \varepsilon$. We now observe that H'^T has entries which are iid copies of \mathcal{X}_p . Thus from Theorem 1, we have $\sigma_{\max}(H'^T) \leq \gamma(\sqrt{r} + \sqrt{m})$ whp., and thus $\|H'^T x\| < 4\gamma\sqrt{n}\|x\|$.

Now we ask: how many eigenvalues of $(\lambda I - D)^{-1}$ have magnitude $> \frac{C}{2\gamma\sqrt{n}}$? (for some parameter C). It is precisely the number of eigenvalues of D which are in a $2\gamma\sqrt{n}/C$ interval around λ . By the semicircle law (Fact 3, in a scale roughly $1/C$), this number is at most $4n/C$ (recall that we assumed $\delta < 1/2$, and hence $r > n/2$). We set $C = 4n/d$.

Let us now recall that $\|\eta\| \geq \varepsilon$, and further we saw that $\|H'^T x\| \leq 4\gamma\sqrt{n}\|x\| < 2\gamma\sqrt{n} \cdot \frac{\varepsilon^2 d}{n}$. Thus $(\lambda I - D)^{-1}$ stretches $H'^T x$ by a factor at least $\frac{n}{d\varepsilon} \cdot \frac{1}{2\gamma\sqrt{n}}$. Thus by Lemma 4 (in the hypothesis, the vector is stretched by a factor τ/ε), we have $\eta = \eta' + z$, where $\|z\| < \varepsilon$, and η' is in the span of eigenvectors of $(\lambda I - D)^{-1}$ of eigenvalues with magnitude at least $\frac{n}{d} \cdot \frac{1}{2\gamma\sqrt{n}}$. The number of such eigenvectors, as we saw above, is precisely d . This completes the proof. \square

We will use this lemma with $\varepsilon = (p\delta)^{1/2}/20$, and $d = \frac{mp\gamma^2}{20 \log(1/\varepsilon)} \approx \frac{\delta np^2}{40 \log(1/\varepsilon)}$, for which we need $\|x\| = \frac{\varepsilon^2 d}{2n} = \frac{\delta^2 p^3}{C \log(1/p\delta)}$, for some absolute constant C .

3.3 Theorem 2: The ε -net argument, finale

In Section 3.2, we saw that we could restrict y to vectors of the kind $s + z$, where s is a unit vector in the span of a set of d vectors (we can assume s is unit even though we did not technically prove it, since y is extremely close to a unit vector if $\|x\|$ is small), for some “small” d , and $\|z\| \leq \varepsilon$. Let us denote the set of all such y by \mathcal{Y} .

Modified ε -Net. We will need an ε -net with additional property. In particular, we need a set N_ε s.t. for every $y \in \mathcal{Y}$, we have $y = y' + \eta$, where $y' \in N_\varepsilon$, $\|\eta\| \leq \varepsilon$, and further $|\langle \eta, \mathbf{1}_r \rangle| \leq \varepsilon/r$. The additional requirement is to handle the fact that we deal with \mathcal{D}_p and not \mathcal{X}_p .

We can find an N_ε of this kind as follows: first we consider a “usual” ε -net, so we have a set M_ε , such that $\mathcal{Y} \subseteq \bigcup_{x \in M_\varepsilon} B(x, \varepsilon)$. Now we can chop up $B(x, \varepsilon)$

for each ε into $2r$ ‘‘slices’’ where for any u, u' in one slice, $\langle u, \mathbf{1}_r \rangle$ and $\langle u', \mathbf{1}_r \rangle$ differ by at most ε/r , and pick the ‘center’ of each slice (the number of slices is $2r$ because in $B(x, \varepsilon)$, the dot-product with $\mathbf{1}_r$ can vary by at most 2ε). Now we use Lemma 2 to prove:

Lemma 5. *Let u be a unit vector in \mathfrak{R}^m satisfying $\|u - \mathbf{1}_m\|^2 < 1$, and H be a matrix in $\mathfrak{R}^{m \times r}$ whose entries are i.i.d. copies of \mathcal{D}_p . Then w.p. $\geq 1 - \frac{1}{\text{poly}(n)}$, we have*

$$\|Hy\|^2 - \langle u, Hy \rangle^2 \geq mp\gamma^2 \|y\|^2 / 10 \quad \text{for all } y \in \mathcal{Y}$$

Proof. Let $\Pi : \mathfrak{R}^m \mapsto \mathfrak{R}^m$ be the operator which projects to the (dimension $m - 1$) subspace orthogonal to u . Now let $\varepsilon := (p\delta)^{1/2}/20$, and N_ε be an ε -net for \mathcal{Y} in the modified sense as above. Now for each $y \in N_\varepsilon$, we have by Lemma 2,

$$\|\Pi Hy\|^2 \geq mp\gamma^2 \|y\|^2 / 10 \quad \text{w.p.} \geq 1 - e^{-mp\gamma^2 / 10}.$$

By our choice of parameters (i.e., d , and because $\delta np^2 \gg \log n$), we have $r \cdot \left(\frac{1}{\varepsilon}\right)^d < \exp(mp\gamma^2/10)$, we have that w.p. $\geq 1 - \frac{1}{\text{poly}(n)}$, $\|\Pi Hy\|^2 \geq mp\gamma^2 \|y\|^2 / 10$ for all $y \in N_\varepsilon$. In particular, since $\|y\| > 3/4$ for all $y \in \mathcal{Y}$, we have $\|\Pi Hy\| \geq \gamma\sqrt{mp}/6$ for all $y \in N_\varepsilon$ w.h.p.

Now suppose $\eta \in \mathfrak{R}^r$, with the property that $\|\eta\| < \varepsilon$, and $|\langle \eta, \mathbf{1}_r \rangle| < \varepsilon/r$. Write $\eta = \eta' + \rho\mathbf{1}_r$, where $\langle \eta', \mathbf{1}_r \rangle = 0$. So we have $|\rho| < \varepsilon/r$. Thus

$$\|H\eta\| = \|H'\eta' + p(mr)^{1/2}\rho\mathbf{1}_m\| \leq \varepsilon\sigma_{\max}(H') + \gamma\sqrt{m}\varepsilon.$$

From section 2, we have $\sigma_{\max}(H') \leq \gamma(\sqrt{m} + \sqrt{r})$ w.h.p. Combining this with the above, we have that

$$\|\Pi Hy\| \geq \gamma\sqrt{mp}/6 - 2\gamma(\sqrt{m} + \sqrt{r})\varepsilon \quad \text{for all } y \in \mathcal{Y}, \text{ whp.}$$

This completes the proof, noting $\varepsilon < (p\delta)^{1/2}/20$.

Lemma 6. *Let B be the adjacency matrix of a $G(m, p)$ random graph, and let H be an $m \times r$ matrix with entries picked iid from \mathcal{D}_p . Then whp. (over the choice of H, B), we have*

$$\|(\lambda I - B)^{-1}Hy\| \geq \frac{\sqrt{p}}{24} \quad \text{for all } y \in \mathcal{Y}.$$

See Appendix A.4 for the proof. This completes the three step proof of Theorem 2.

4 Nodal domains

Let $G \sim G(n, p)$, and let us consider an eigenvector v of the adjacency matrix A . Let the corresponding eigenvalue be λ , and that it is not the largest (this means $\langle v, \mathbf{1}_n \rangle$ is *small*, and hence we expect it to have both positive and negative

components). Recall that nodal domains of v are maximal connected components of v which have the same sign. For technical reasons, we will consider only *weak* nodal domains, i.e., if a coordinate $v_i = 0$, we allow it to be in multiple domains. In fact, Courant’s nodal domain theorem in a discrete setting applies only to weak nodal domains (though there is a different version for *strong* nodal domains, it is not as clean to state) [9].

Intuitively if a graph is fairly dense, we would expect much fewer than n nodal domains for any function f on the vertices, unless the graph has a certain structure (for instance if a graph is bipartite, there exists an f with n different nodal domains). Thus for *dense enough* random graphs, we would expect to see precisely two nodal domains (one each for positive/negative) when we consider the eigenfunctions. This will be the subject of this section.

We crucially use the following (very recent) result of Erdős, Knowles, Yau and Yin [10]. A weaker result (used in an earlier version of the current paper) was proved by Tran *et al* [11].

Lemma 7. *[[10], Theorem 2.16] Let A be the adjacency matrix of $G(n, p)$, and let \mathbf{v} be any eigenvector of A , normalized to $\|\mathbf{v}\|_2 = 1$. Then we have*

$$\|\mathbf{v}\|_\infty \leq \frac{\log^2 n}{\sqrt{n}}, \quad \text{whp.}$$

For convenience, we will denote this upper bound by β . Let us start with some simple observations, along the lines of [6]. First, observe that the size of the largest independent set in G is $O(\log n)/p$ whp. (By a standard calculation – Observation 1). Thus there cannot be more than $O(\log n)/p$ domains of each sign (because vertices in different domains cannot have an edge by definition). Further there can be at most one domain of each sign of size $\geq C \log n/p$, because otherwise we have two sets of size $C \log n/p$ with no edges between them and this cannot exist whp (calculation à la Observation 1). Thus let M_+ and M_- denote the largest positive and negative domains respectively. We claim that both are ‘large’ whp.

Lemma 8. *Suppose $p > n^{-1/2+\varepsilon}$. Then both $|M_+|$ and $|M_-|$ are at least $\log^2 n/p$ whp.*

The proof is an application of Lemma 7, and can be found in Section A.5. Now, following [6], we will call the vertices in $M_e = V \setminus (M_+ \cup M_-)$ *exceptional vertices*. At most $C \log n/p$ of these have $v_i > 0$, because otherwise these along with M_+ are two large sets with no edges between them. Thus the number of exceptional vertices is only $C \log n/p$. This crude upper bound will be helpful in showing that there are in fact no exceptional vertices w.h.p.

Theorem 3. *Let $p \geq n^{-1/19+\varepsilon}$, and v be an eigenvector of $G \sim G(n, p)$ for non-first eigenvalue λ . Then v has precisely two nodal domains whp.*

Proof. Equivalently, we wish to prove that there are no exceptional vertices. Suppose for the sake of contradiction that $M_e \neq \emptyset$, and let i be an exceptional

vertex. Wlog, we may assume that $v_i > 0$. Thus i can have edges only to $M_- \cup M_e$, because if it has an edge to M_+ it is in the same connected component and hence not exceptional. Consider the i th coordinate in the equality $Av = \lambda v$

$$\sum_{j \in M_e \cap \Gamma(i)} v_j + \sum_{j \in M_- \cap \Gamma(i)} v_j = \lambda v_i, \quad \text{i.e.,} \quad \sum_{j \in M_e \cap \Gamma(i)} v_j - \sum_{j \in M_- \cap \Gamma(i)} |v_j| = \lambda v_i \quad (3)$$

The proof is along the following lines: for every vertex i , we wish to claim that $\sum_{j \in \Gamma(i)} |v_j|$ is large. This is because Theorem 2 implies $\sum_{j \in \Gamma(i)} v_j^2$ is ‘large’ (because the ℓ_2 mass on any set is large, and we apply this to the set of neighbors), and then using Lemma 7. Then we will claim that the other two terms are both very small – the first is because there are too few terms, and the last by Lemma 7 and the fact that λ is a non-first eigenvalue.

Formally, we first note that $|\Gamma(i)| \geq np(1 - o(1))$ for every i whp. Thus by Theorem 2 applied to $\Gamma(i)$,³ $\sum_{j \in \Gamma(i)} v_j^2 \geq s$ whp., where s denotes the RHS of the bound in Theorem 2. We can then take a union bound and conclude this for all i . Now by Lemma 7 we have $|v_j| \leq \beta$ for every j , thus $\sum_{j \in \Gamma(i)} |v_j| \geq s/\beta$. We showed that $|M_e|$ is at most $\frac{O(\log n)}{p}$, and thus

$$\sum_{j \in M_e \cap \Gamma(i)} |x_j| \leq \frac{C\beta \log n}{p},$$

for some constant C . We will verify that our choice of parameters satisfy

$$\frac{s}{2\beta} \geq \frac{4C\beta \log n}{p}, \quad \text{and} \quad \frac{s}{2\beta} > 8\gamma\sqrt{n} \geq 4\lambda\beta. \quad (4)$$

These will together contradict (3), thus finishing the proof of the theorem.

Plugging in the bound from Theorem 2 with $\delta = p$, the first inequality above simplifies to $\frac{p^{10}}{\log(1/p)} \geq \frac{4C\beta \log n}{np}$, which is true as long as $p \geq n^{-1/11+\epsilon}$. The second inequality in (4), along with the fact that $\lambda \leq 2\gamma\sqrt{n} = 2\sqrt{np(1-p)}$ simplifies to $\frac{p^{10}}{\log^2(1/p)} \geq \frac{(np)^{1/2} \log^4 n}{n}$, which is true as long as $p \geq n^{-1/19+\epsilon}$. The latter bound dominates.

5 Future Directions

A natural question posed by [11] is the following

³ Strictly speaking, we cannot use the theorem – because it says that for a fixed S , the property holds whp. In our case we have the graph, and we are picking $S = \Gamma(i)$. However, we can repeat the proof of the lemma setting $S = \Gamma(i)$ for one vertex i – in the block decomposition, one of the columns of H has all ones, and the ‘corresponding’ column in D has zeroes. However the rest of the entries are random, and this ensures that all the claims hold. We omit the details.

Conjecture 1. Let v be an eigenvector corresponding to a non-first eigenvalue of A , and let u be any unit vector. Then whp., $|\langle u, v \rangle| \leq \frac{C \log n}{\sqrt{n}}$.

Note that ℓ_∞ bounds on eigenvectors translate to choosing $u = e_i$. While it seems tempting to use similar techniques – after all, by rotation we can change any u to e_i – the difficulty is that we no longer have *blocks* with entries that are independent of each other.

Level sets. One can also consider nodal domains of *level sets* of the eigenvectors: these are nodal domains of the vector $v - t\mathbf{1}_n$, for some threshold t . Such sets arise in spectral partitioning, for instance [16]. We would expect that for t beyond some critical value, the nodal domains start “breaking apart” into smaller pieces, and it is interesting to study this threshold.

$G(n, p)$ for smaller p . Perhaps the most interesting open question is to study delocalization and properties of nodal domains for small p (including $p = C/n$ for large constant C). Many of our technical ingredients, like the delocalization results of Erdős, Schlein and Yau, and even some properties about the spectrum are not known for this range of p . An ambitious goal is to investigate nodal domain counts for the giant component. Perhaps a more approachable question is to reduce the lower bound on p required by Theorem 3 (i.e., does it hold for all $p > \omega(\log n)/n$?)

5.1 Acknowledgements

We thank Peter Sarnak for pointing us to [5] which served as the starting point for this work. We would also like to thank James Lee and Nati Linial for interesting discussions, and Nati for encouraging thought on related problems, such as nodal domains in level sets.

References

1. Wigner, E.: On the Distribution of the Roots of Certain Symmetric Matrices. *Annals of Mathematics* **67** (1958) 325–328
2. Bollobas, B.: *Random Graphs*. Cambridge University Press (2001)
3. Erdos, L.: Universality of Wigner random matrices: a Survey of Recent Results. ArXiv e-prints (April 2010)
4. Tao, T., Vu, V.: Random matrices: Universality of local eigenvalue statistics. ArXiv e-prints (June 2009)
5. Erdős, L., Schlein, B., Yau, H.T.: Local semicircle law and complete delocalization for wigner random matrices. *Communications in Mathematical Physics* **287** (2009) 641–655 10.1007/s00220-008-0636-9.
6. Dekel, Y., Lee, J.R., Linial, N.: Eigenvectors of random graphs: Nodal domains. In: APPROX-RANDOM. (2007) 436–448
7. : Nodal patterns in physics and mathematics. *The European Physical Journal - Special Topics* **145** (June 2007) 10.1140/epjst/e2007-00142-7.

8. Band, R., Oren, I., Smilansky, U.: Nodal domains on graphs - How to count them and why? ArXiv e-prints (November 2007)
9. Bıykođlu, T., Leydold, J., Stadler, P.F.: Laplacian eigenvectors of graphs: Perron-Frobenius and Faber-Krahn type theorems. Springer Verlag (2007)
10. Erdos, L., Knowles, A., Yau, H.T., Yin, J.: Spectral Statistics of Erd\H{o}s-R'enyi Graphs I: Local Semicircle Law. ArXiv e-prints (March 2011)
11. Tran, L., Vu, V., Wang, K.: Sparse random graphs: Eigenvalues and Eigenvectors. ArXiv e-prints (November 2010)
12. Tao, T.: Topics in random matrix theory (in preparation). (2011)
13. Mitra, P.: Entrywise bounds for eigenvectors of random graphs. Electronic Journal of Combinatorics **16** (2009)
14. Bai, Z., Yin, Y.: Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. Ann. Probab. **21**(3) (1993) 1275–1294
15. Rudelson, M., Vershynin, R.: Smallest singular value of a random rectangular matrix. Communications on Pure and Applied Mathematics **62**(12) (2009) 1707–1739
16. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **22** (1997) 888–905

A Additional Lemmas

Lemma 9. *Let Y_1, \dots, Y_n be i.i.d. random variables drawn according to \mathcal{X}_p , and let $\alpha_1, \dots, \alpha_n$ be real numbers s.t. $\sum_i \alpha_i^2 = S$. Further let Z denote the r.v. $Z := \sum_i \alpha_i Y_i$. Then*

$$\begin{aligned}\mathbb{E}[Z^2] &= Sp(1-p) = S\gamma^2 \\ \mathbb{E}[Z^4] &\leq S^2(\gamma^2 + 3\gamma^4)\end{aligned}\tag{5}$$

Proof. Noting that Y_i are drawn from the symmetric distribution \mathcal{X}_p , we have that $\mathbb{E}[Y_i Y_j] = \mathbb{E}[Y_i^3 Y_j] = 0$ for distinct i, j . Thus we have

$$\begin{aligned}\mathbb{E}[Z^2] &= \mathbb{E}\left[\left(\sum_i \alpha_i Y_i\right)^2\right] = \mathbb{E}\left[\sum_i \alpha_i^2 Y_i^2\right] = \gamma^2 \cdot S \\ \mathbb{E}[Z^4] &= \mathbb{E}\left[\left(\sum_i \alpha_i Y_i\right)^4\right] = \mathbb{E}\left[\sum_{i,j,k,l} \alpha_i \alpha_j \alpha_k \alpha_l Y_i Y_j Y_k Y_l\right] \\ &= \sum_i \alpha_i^4 \cdot \mathbb{E}[Y_i^4] + 6 \sum_{i < j} \alpha_i^2 \alpha_j^2 \cdot \mathbb{E}[Y_i^2 Y_j^2]\end{aligned}$$

Now $\mathbb{E}[Y_i^4] = p(1-p)(1-3p+3p^2) < p(1-p) = \gamma^2$, and $\mathbb{E}[Y_i^2 Y_j^2] = \gamma^4$. Noting that $\sum_i \alpha_i^4 < S^2$ and $6 \sum_{i < j} \alpha_i^2 \alpha_j^2 < 3S^2$ now implies the desired result.

Observation 1 *Let $G \sim G(n, p)$. The size of the largest independent set in G is at most $C \log n/p$ whp.*

Proof. The expected number of independent sets of size k is (by linearity of expectation):

$$\binom{n}{k} e^{-k^2/2} \leq e^{-k(\log n - (k/2))}.$$

This number is $\ll 1/\text{poly}(n)$ for $k > C \log n/p$, thus Markov's inequality gives the result.

A.1 Proof of Claim. A.2

Claim. Let $y \in \mathfrak{R}^r$ be any vector with $\|y\| = 1$. Let H' be an $m \times r$ matrix with entries drawn i.i.d. from \mathcal{X}_p , and let $Z = H'y$. Then for any unit vector $u \in \mathfrak{R}^m$, and all $t > 0$, we have

$$\Pr [|\langle u, Z \rangle| \geq t] \leq e^{-t^2}. \quad (6)$$

Proof. We can write out $u^T H'y = \sum_{i,j} H'_{i,j} u_i y_j$. Since $\sum_{i,j} u_i^2 y_j^2 = 1$, and H' are mean-zero random variables bounded in absolute value by 1, we have the claim by Azuma's inequality.

A.2 Proof of Lemma 2

First, let us write $H = H' + pJ_{m,r}$, as before. Also since the statement is invariant under scaling y , let us assume $\|y\| = 1$. Further, let $y = y' + \alpha \mathbf{1}_r$, where $\langle y', \mathbf{1}_r \rangle = 0$. Then $Hy = H'y + (mr)^{1/2} p\alpha \cdot \mathbf{1}_m$. Let us write $c := (mr)^{1/2} p\alpha$, and let $Z = H'y$. Thus for random H , Z is a vector with i.i.d. mean-zero entries. The quantity we are interested in is now

$$\begin{aligned} & \|Z + c\mathbf{1}_m\|^2 - \langle u, (Z + c\mathbf{1}_m) \rangle^2 \\ &= \|Z\|^2 + \|c\mathbf{1}_m\|^2 + 2\langle Z, c\mathbf{1}_m \rangle - (\langle u, Z \rangle^2 + \langle u, c\mathbf{1}_m \rangle^2 + 2\langle u, Z \rangle \langle u, c\mathbf{1}_m \rangle) \\ &= \left(\|Z\|^2 - \langle u, Z \rangle^2 \right) + \left(\langle \mathbf{1}_m, c\mathbf{1}_m \rangle^2 - \langle u, c\mathbf{1}_m \rangle^2 + 2\langle \mathbf{1}_m, Z \rangle \langle \mathbf{1}_m, c\mathbf{1}_m \rangle - 2\langle u, Z \rangle \langle u, c\mathbf{1}_m \rangle \right) \end{aligned}$$

Let \mathcal{T}_1 and \mathcal{T}_2 denote the terms in the first and the second parentheses. Consider \mathcal{T}_2 . To simplify things, let $\delta := \mathbf{1}_m - u$, so that $\mathbf{1}_m = u + \delta$. So first, we have

$$\langle \mathbf{1}_m, c\mathbf{1}_m \rangle^2 - \langle u, c\mathbf{1}_m \rangle^2 = \langle \delta, c\mathbf{1}_m \rangle \langle \mathbf{1}_m + u, c\mathbf{1}_m \rangle.$$

Next, writing $\langle \mathbf{1}_m, Z \rangle = \langle u, Z \rangle + \langle \delta, Z \rangle$,

$$\langle \mathbf{1}_m, Z \rangle \langle \mathbf{1}_m, c\mathbf{1}_m \rangle - \langle u, Z \rangle \langle u, c\mathbf{1}_m \rangle = \langle \delta, Z \rangle \langle \mathbf{1}_m, c\mathbf{1}_m \rangle + \langle u, Z \rangle \langle \delta, c\mathbf{1}_m \rangle.$$

Now since u and $\mathbf{1}_m$ are unit vectors, and $u = \mathbf{1}_m - \delta$, we have $1 = 1 + \|\delta\|^2 - 2\langle \delta, \mathbf{1}_m \rangle$, implying $\langle \delta, \mathbf{1}_m \rangle = \|\delta\|^2/2$. Thus $\langle \mathbf{1}_m + u, c\mathbf{1}_m \rangle = 2c(1 - \|\delta\|^2/4)$. Plugging in these values and simplifying,

$$\mathcal{T}_2 = c^2 \|\delta\|^2 \cdot (1 - \|\delta\|^2/4) + 2c(\langle \delta, Z \rangle + \frac{\|\delta\|^2}{2} \langle u, Z \rangle)$$

By assumption, $\|u - \mathbf{1}_m\|^2 < 1$, and so we have

$$\mathcal{T}_2 \geq \frac{c^2 \|\delta\|^2}{2} - 2|c|(|\langle \delta, Z \rangle| + \frac{\|\delta\|^2}{2} |\langle u, Z \rangle|).$$

Now a standard Azuma-Hoeffding bound gives the following.

Claim. Let $y \in \mathfrak{R}^r$ be any vector with $\|y\| = 1$. Let H' be an $m \times r$ matrix with entries drawn i.i.d. from \mathcal{X}_p , and let $Z = H'y$. Then for any unit vector $u \in \mathfrak{R}^m$, and all $t > 0$, we have

$$\Pr [|\langle u, Z \rangle| \geq t] \leq e^{-t^2}. \quad (7)$$

The simple proof can be found in Section A.1.⁴ Using this we have that $\Pr[|\langle \delta, Z \rangle| > \|\delta\| \cdot \gamma\sqrt{mp}/20] < e^{-mp\gamma^2/400}$, and so also $\Pr[|\langle u, Z \rangle| > \gamma\sqrt{mp}/20] < e^{-mp\gamma^2/400}$. Thus w.p. at least $1 - e^{-mp\gamma^2/500}$, we have

$$\mathcal{T}_2 \geq \left(\frac{(c\|\delta\|)^2}{4} - 2|c|\|\delta\| \cdot \frac{\gamma\sqrt{mp}}{20} \right) + \|\delta\|^2 \left(\frac{c^2}{4} - 2|c| \cdot \frac{\gamma\sqrt{mp}}{20} \right)$$

Since $\|\delta\|^2 < 2$, and for all $\alpha, \beta \in \mathfrak{R}$ we have $\alpha^2 - 2\alpha\beta \geq -\beta^2$, we have, w.p. at least $1 - e^{-mp\gamma^2}$,

$$\mathcal{T}_2 \geq -mp\gamma^2/50$$

Now consider $\mathcal{T}_1 := \|Z\|^2 - \langle u, Z \rangle^2$. As before, we have that $\Pr[\|\langle u, Z \rangle\|^2 > mp\gamma^2/100] < e^{-mp\gamma^2/100}$. We now show that $\|Z\|^2 \geq mp\gamma^2/10$ with extremely high probability. More specifically,

Claim. Let $y \in \mathfrak{R}^r$ be given, with $\|y\|^2 = 1$, and H' be an $m \times r$ matrix with entries drawn i.i.d. from \mathcal{X}_p . Then

$$\Pr \left[\|H'y\|^2 < \frac{mp\gamma^2}{10} \right] < e^{-mp/10}.$$

Proof. Let $Z := H'y$, and let Z_i denote the i th component of Z . We can check (see Lemma 9 for details) that for each i ,

$$\mathbb{E}[Z_i^2] = \gamma^2\|y\|^2 \quad (8)$$

$$\mathbb{E}[Z_i^4] \leq (\gamma^2 + 3\gamma^4)\|y\|^4 \quad (9)$$

Thus by the Paley-Zygmund inequality, $\Pr[Z_i^2 > \gamma^2\|y\|^2/3] \geq \frac{1}{3} \cdot \frac{(\gamma^2\|y\|^2)^2}{(\gamma^2 + 3\gamma^4)\|y\|^4} \geq p/6$. [The last bit is because we assume $p \leq 1/2$.] Let us say an index i is *good* if the above inequality holds. Thus any i is good w.p. at least $p/6$, and further, these events are independent. Thus in expectation there are at least $mp/6$ good indices, and by simple Chernoff bounds, the probability that there are at most $mp/8$ is at most $e^{-mp/100}$. If there are at least $mp/8$ good indices, then $\|Z\|^2 \geq (mp/8) \cdot \gamma^2/3 \geq mp\gamma^2/24$. This proves the claim.

Putting all these together, we showed that $\mathcal{T}_1 \geq \frac{mp\gamma^2}{10}$ w.p. at least $1 - e^{-mp\gamma^2/20}$. Also, we saw that $\mathcal{T}_2 \geq -\frac{mp\gamma^2}{20}$ w.p. at least $1 - e^{-mp\gamma^2/20}$. This completes the proof of Lemma 2.

⁴ It would be nice if we could have $\exp(-t^2/2\gamma^2)$ in the RHS – it would lead to a better bound overall.

A.3 Proof of Lemma 4

Proof. Let $\lambda_1, \lambda_2, \dots, \lambda_n$ be the eigenvalues of M in decreasing order of magnitude, and let s denote the index where the values become smaller than τ . Further, let w_i be the corresponding eigenvectors, and let $u = \sum_i \alpha_i w_i$. Then $Mu = \sum_i \alpha_i \lambda_i w_i$, and thus if we write $v = v' + z$, with v' being the component along the ‘large’ eigenvalues of M , we have

$$\|z\|^2 < \sum_{i>s} \lambda_i^2 \alpha_i^2 < \tau^2 \sum_{i>s} \alpha_i^2 < \varepsilon^2$$

□

A.4 Proof of Lemma 6

Proof. Let u be the eigenvector corresponding to the largest eigenvalue of B . It is well-known (see e.g. [6]) that u is “close” to $\mathbf{1}_m$, i.e., $\|u - \mathbf{1}_m\|^2 < 1$ whp. Now from the semicircle law (section 2), all the eigenvalues of B except the first are at most $3\gamma\sqrt{m}$ whp. Thus if w is a vector orthogonal to u , we have that $\|(\lambda I - B)^{-1}w\| \geq \frac{1}{3\gamma\sqrt{m}} \cdot \|w\|$. Now using the previous lemma, we obtain

$$\|(\lambda I - B)^{-1}Hy\| \geq \frac{1}{3\gamma\sqrt{m}} \cdot \frac{\gamma\sqrt{mp}}{8} \geq \frac{\sqrt{p}}{24} \quad \text{for all } y \in \mathcal{Y}.$$

□

A.5 Proof of Lemma 8

Proof. Suppose for the sake of contradiction, that $|M_-| \leq \log^2 n/p$. By the above, there can only be \log^2/p^2 indices i with $v_i < 0$ (small number of components, each small).

Since λ is not the top eigenvalue, $\langle v, \mathbf{1}_n \rangle \leq \frac{C \log n}{\sqrt{np}}$ whp. (c.f. [6]), i.e., $\sum_i v_i \leq \frac{C \log n}{\sqrt{p}}$. But using Lemma 7, we get $\sum_i |v_i| \geq \frac{1}{\beta} \sum_i v_i^2 = \frac{1}{\beta}$. Since the sum is small, the negative terms should cancel the positive ones, however there are only $\log^2 n/p^2$ of them and each is of magnitude only β . Thus this would be a contradiction if

$$\beta \cdot \frac{\log^2 n}{p^2} < \frac{1}{4\beta}.$$

This simplifies to $p^2 > \log^6 n/n$, which holds as long as $p > n^{-1/2+\varepsilon}$.