

THE LIMITING SPECTRAL LAW FOR SPARSE IID MATRICES

ASHWIN SAH, JULIAN SAHASRABUDHE, AND MEHTAAB SAWHNEY

ABSTRACT. Let A be an $n \times n$ matrix with iid entries where $A_{ij} \sim \text{Ber}(p)$ is a Bernoulli random variable with parameter $p = d/n$. We show that the empirical measure of the eigenvalues converges, in probability, to a deterministic distribution as $n \rightarrow \infty$. This essentially resolves a long line of work to determine the spectral laws of iid matrices and is the first known example for non-Hermitian random matrices at this level of sparsity.

1. INTRODUCTION

For an $n \times n$ matrix M , define its *spectral distribution* to be the probability measure on \mathbb{C} , which puts a point mass of equal weight on each eigenvalue λ of M :

$$\mu_M = n^{-1} \sum \delta_\lambda.$$

One of the central projects in random matrix theory, going back to the seminal 1958 work of Wigner [34], is to determine the limiting spectral distribution of various random matrix models as the dimension tends to infinity.

While this area has enjoyed spectacular advances in the 80 years since its inception, several fundamental matrix models have eluded all attempts to understand their spectral law. Two major problems here concern very sparse matrices, in particular matrices with a constant number of non-zero entries in a typical row or column. The first is to show that μ_{M_n} tends to the oriented Kesten–McKay law as $n \rightarrow \infty$ when M_n is an $n \times n$ matrix chosen uniformly at random from all matrices with exactly $d \in \mathbb{N}$ ones in each row and column (so called d -regular digraphs). The second is to show the existence of the limiting spectral distribution for iid Bernoulli random matrices with parameter $p = d/n$, for d fixed.

In this paper we resolve this latter conjecture. As will see, this is the last piece in a complete understanding of the limiting spectral laws of iid random matrices and is the first time the existence of a limit law has been established for *any* non-Hermitian random matrix model at this level of sparsity. In particular, this resolves a question highlighted by Tikhomirov in his 2022 ICM survey [32, Problem 6].

Theorem 1.1. *For $d > 0$, and each n , let A_n be an $n \times n$ matrix with iid entries distributed as $\text{Ber}(d/n)$. There exists a distribution μ_d on \mathbb{C} so that μ_{A_n} converges to μ_d , in probability.*

Our proof differs significantly from previous approaches, such as [25], and, for example, entirely avoids the direct use ε -nets. Rather, our approach is to “build up” the matrix, a row and column at a time, and study the evolution of the point process defined by the singular values of the shifted matrices $A_n - zI$ as we add rows and columns. Our methods additionally give a considerably shorter proof of the difficult and celebrated theorem of Rudelson and Tikhomirov [25] who proved the existence of the limiting spectral law in the case $pn \rightarrow \infty$. The details of this are contained in the sister paper [27].

We remark that the real problem here is for $d > 1$. In the “subcritical” and “critical” regimes, $d < 1$ and $d = 1$, it is not hard to show that almost all of the eigenvalues of A_n are

Sah was supported by the PD Soros Fellowship. Sah and Sawhney were supported by NSF Graduate Research Fellowship Program DGE-2141064. A portion of this work was conducted when Sawhney was visiting Cambridge with support from the Churchill Scholarship.

0 and therefore $\mu_d = \delta_0$. We include these details in Section 3. On the other hand, we expect that for $d > 1$, μ_d is a rich, non-trivial distribution. Thus $p = 1/n$ represents the threshold for the “birth” of the spectrum of A_n .

While in this paper we are almost exclusively focused on the *existence* of the limiting distribution μ_d , many properties of μ_d can be deduced, using other methods, now that it has been shown to exist. For example, it is possible to show that each μ_d is rotationally invariant, which incidentally falls out of some calculations we need in the course of proving Theorem 1.1 (see Lemma 11.1).

Let us also remark that our proof also can be adapted to the case where all non-zero entries are iid copies of a random variable ξ with variance 1 and with moments that decay sufficiently quickly.

So far we have been somewhat loose in our discussion of the precise mode of convergence for the random measures μ_{A_n} . In Theorem 1.1, and throughout this paper, we are concerned with convergence *in probability*: a sequence of random measures μ_n converges *in probability* to a probability measure μ , if for all continuous bounded functions $f: \mathbb{C} \rightarrow \mathbb{C}$, and all $\varepsilon > 0$, we have

$$(1) \quad \left| \int f d\mu_n - \int f d\mu \right| < \varepsilon$$

with probability $1 - o(1)$. If this holds we write $\mu_n \rightsquigarrow \mu$. It is also natural to consider the stronger notion of *almost sure convergence* of μ_{A_n} to μ_d , which is a problem we leave open for future work.

1.1. The least singular value problem. Before we discuss the history of the limiting spectral laws for iid random matrices, we highlight a consequence of our results that is of independent interest and essentially resolves another question raised by Tikhomirov in his ICM survey [32, Problem 7].

Here we are interested in proving that the spectrum of A_n does not “clump” about a point $z \in \mathbb{C}$. This “clumping” is captured in the extreme behaviour of the least singular value of the random *shifted* matrices $A_n - zI$. Recall that for an $n \times n$ matrix M , its least singular value is

$$\sigma_n(M) = \min_{v \in \mathcal{S}^{n-1}} \|Mv\|_2.$$

In this paper we prove the following “qualitative” estimates on $\sigma_n(A_n - zI)$ conjectured by Tikhomirov [32].

Theorem 1.2. *Fix $d > 1$ and $\varepsilon > 0$. Then for Lebesgue almost all z we have the following. For each n , let A_n be an $n \times n$ matrix with iid entries distributed as $\text{Ber}(d/n)$. Then*

$$\mathbb{P}\left(\sigma_n(A_n - zI) \leq \exp(-\varepsilon n)\right) = o(1).$$

In fact, a careful analysis of our proof reveals that we may take $\varepsilon = n^{-1/2+o(1)}$ and the $o(1)$ probability bound can be taken to be $(\log n)^{-\Omega(1)}$. However, these quantitative aspects are not the focus of this work and thus our methods are not tailored to this problem. We briefly discuss these dependencies in Remark 13.6.

We also note that the proof in the paper can be modified to handle $d < 1$ in the setting of Theorem 1.2, but we do not pursue the study of this sub-critical regime here, in the interest of brevity.

1.2. History of the limiting spectral law for iid random matrices. The project of determining the limit laws for random matrix models goes back to the seminal work of Wigner who proved the famous “semi-circular” law for random *symmetric* matrices (or Wigner matrices). We let M_n be an $n \times n$ random symmetric matrix with entries $(M_n)_{i \leq j}$ uniform in $\{0, 1\}$. Since symmetric matrices have all real eigenvalues it then makes sense to define, for $a < b$, $N_n(a, b)$ to be the number of eigenvalues of M_n in the interval (a, b) . Wigner’s semi-circular law says that

$$\lim_{n \rightarrow \infty} \frac{N_n(a\sqrt{n}, b\sqrt{n})}{n} = \frac{1}{2\pi} \int_a^b (4 - x^2)_+^{1/2} dx$$

almost surely.

While these methods led to a very good understanding of symmetric and Hermitian random matrix models, determining the limiting spectral distribution for matrices with *iid* (non-symmetric) entries proved to be substantially more difficult. Here the first steps were taken by Mehta [22], who in the 1960s showed that when A_n has iid complex Gaussian entries, the spectral distribution of $n^{-1/2}A_n$ converges to the uniform measure on the unit disc $\{z \in \mathbb{C}: |z| \leq 1\}$, the so-called *circular law*. Mehta’s proof relied deeply on the symmetries of complex Gaussian random variables and it was not until the 1990s that Edelman [9] managed to prove the same result for *real* Gaussian random variables.

The case of more general coefficient distributions was studied in the 1980s by Girko [12], who developed very influential ideas such as the “hermitization” technique, but his method relied on an unproven statement about the least singular value of iid matrices. This statement was then circumvented in the 1990s by Bai [1], who extended the theorem of Edelman to matrices where the entries are iid mean 0, variance 1, and satisfy some smoothness and moment conditions. These results were then improved by Götze and Tikhomirov [14], Pan and Zhou [23] and Tao and Vu [29], by using the method of Bai along with methods of Rudelson and Vershynin [26] and Tao and Vu [31] to control the least singular value. Finally, Tao and Vu [30] proved the full “universality” theorem, showing the circular law holds for *any* sequence of iid matrix with entries distributed as a mean 0 variance 1 random variable. The method, relying on their breakthroughs in inverse Littlewood–Offord theory, provides a full understanding of empirical spectral distributions of “dense” random matrices.

Theorem 1.3 (Tao and Vu). *Let ξ be a complex random variable with mean 0 and variance 1, let A_n be a sequence of random matrices with iid entries distributed as ξ . If we put $A_n^* = A_n \cdot n^{-1/2}$ then the spectral measure $\nu_{A_n^*}$ converges to the circular law in probability¹.*

While this celebrated line of results gives us a very good understanding of the limiting spectral laws of dense matrices, it does not tell us anything about matrices where the non-zero entries are *sparse*, as is often interesting in combinatorial settings. Of particular interest are Bernoulli random matrices: random iid matrices where all entries are Bernoulli random variables that take 1 with probability $p = p_n \rightarrow 0$ and 0 otherwise.

The spectral laws of such matrices were considered by Götze and Tikhomirov [14], who proved that the limiting spectral law of A_n is still the circular law (with appropriate normalization) for all $p > n^{-1/4+\varepsilon}$. Tao and Vu [29] improved this range to $p > n^{-1+\varepsilon}$, and Basak and Rudelson [2] improved this range further to account for all $p > \omega(n^{-1}(\log n)^2)$.

Then, in an important and difficult paper, Rudelson and Tikhomirov [25] extended these results to account for all $pn \rightarrow \infty$. This work is of particular interest since it is not hard to see that that the condition $pn \rightarrow \infty$ is *necessary* for convergence to the circular law: for pn bounded there is always an atom at zero.

¹In fact, Tao and Vu prove this result for the stronger notion of “almost sure” convergence.

Theorem 1.4 (Rudelson and Tikhomirov). *Let $p = p_n$ be such that $pn \rightarrow \infty$ and $p \rightarrow 0$. For each n , let A_n be an $n \times n$ matrix with iid entries distributed as $\text{Ber}(p)$. If we put $A_n^* = (pn)^{-1/2}A_n$ then $\mu_{A_n^*}$ tends to the circular law, in probability.*

In fact, they prove a more general result that allows for each non-zero entry to be a copy of an iid random variable ξ with mean 0 and variance 1. We point the reader to [25] or our sister paper [27], where we give a simple proof of a variant of this theorem that subsumes all of these previous results.

This leaves open what has proven to be the most difficult and subtle case, the case of $p = d/n$ for constant $d > 0$. As mentioned before, the real problem is for $d > 1$, as it is not too hard to see that² for $d \leq 1$, the limiting measure is the point mass at 0. As soon as $d > 1$, however, the limiting distribution μ_d becomes a rich and interesting distribution. In this paper we establish the existence of the limiting spectral law for all d , down to its “birth” at $d > 1$.

2. OUTLINE OF PROOF

Before describing our method, we note that in our paper [27] we adapt the methods of this paper to reprove the sparse circular law of Rudelson and Tikhomirov [25]. There, we are dealing with matrices for which an entry is nonzero with probability p , where $pn \rightarrow \infty$, which allows us to avoid several significant challenges that occur in this paper. Thus one may find it easier to absorb our method by first understanding [27] and then returning to this paper. Of course, we will not assume any knowledge of [27] in our treatment here.

2.1. Convergence of the logarithmic potential. To establish the convergence of the spectral law, it is enough to prove the point-wise, almost everywhere, convergence of the logarithmic potential of the spectral law, which is (although we don’t need this expression here) the (random) function

$$U_n(z) = -\frac{1}{n} \sum_{\lambda} \log |\lambda - z|,$$

where the sum is over the eigenvalues λ of our random matrix A_n . Note here the limit necessarily does not match logarithmic potential of the circular law, which presents a key difficulty. We start by using Girko’s “hermitization” method (see e.g. [5]) to express

$$(2) \quad U_n(z) = -\frac{1}{n} \sum_{j=1}^n \log (\sigma_j(A_n - zI_n)),$$

where $\sigma_1(M) \geq \dots \geq \sigma_m(M)$ denote the (right) singular values of the $n \times m$ matrix M .

The big advantage of this expression is that it is in terms of *singular values*, rather than eigenvalues, which have the advantage that they are the *real* and, in particular, the eigenvalues of the *Hermitian* matrix $(A_n - zI_n)^\dagger(A_n - zI_n)$. The point is that, if we define the measure

$$\nu_z = n^{-1} \sum \delta_\sigma,$$

where the sum is over the singular values $\sigma \in \{\sigma_n(A_n - zI_n), \dots, \sigma_1(A_n - zI_n)\}$, we can recover the bulk behaviour of ν_z by simply computing the trace moments of $(A_n - zI_n)^\dagger(A_n - zI_n)$, using now standard techniques.

While this is a good (and far from novel) first step, this *does not* imply the convergence of the log potential since it is *a priori* possible for some singular values in the sum (2) to disrupt the convergence of the bulk, either by being very large or by being very small. The possible problem of σ_1 being large is easily brought under control by standard estimates, which brings us to the heart of the matter: abnormally *small* singular values.

²This is also implicit in the work of Coste [6] for $d < 1$.

Traditionally, one needs to prove estimates roughly of the form

$$(3) \quad \mathbb{P}(\sigma_{n-k}(A_n - zI) \leq \exp(-\varepsilon n/k)) = o(1),$$

for any $\varepsilon > 0$, all $k \leq n - 1$, and for Lebesgue almost every $z \in \mathbb{C}$. Indeed, *heuristically* we have $\sigma_{n-k} = \Theta(kd^{1/2}n^{-1})$, typically, and thus (3) perhaps appears to be an easily surmounted obstacle, as it represents (what we expect to be) extremely abnormal behaviour. However, obtaining bounds of this type has recently been *the* significant challenge in this area. Indeed, proving (3) is one of the principle achievements of the work Tao and Vu [30] in their work on the circular law for dense matrices. For sparse matrices, the challenge is greater still as there is less “randomness” to use. For their sparse circular law, Rudelson and Tikhomirov [25] develop a whole toolbox of sophisticated techniques to prove singular value estimates of the type³ (3).

2.2. The evolution of windows of singular values. In this paper, our focus is slightly different, and we instead look to control the bottom *window* of singular values

$$(4) \quad W_{n,0}(z) = -\frac{1}{n} \sum_{j=0}^{\delta n} \log(\sigma_{n-j}(A_n - zI_n)),$$

by showing that for all $\varepsilon > 0$ and almost all $z \in \mathbb{C}$, we have that

$$(5) \quad \mathbb{P}(W_{n,0}(z) \leq \varepsilon) = 1 - o(1).$$

Here δ is chosen to be sufficiently small relative to d, z, ε .

While this, so far, is not much of a departure from the task of proving (3), our principal difference comes from how we approach $W_{n,0}(z)$, which is fundamentally dynamic: we build up the randomness in the matrix bit by bit and track how the singular values evolve.

To best explain this, we first describe an approach that is simpler than our true approach but only works for sufficiently large (but fixed) d . We then describe a more complicated revelation process that works for all d down to the threshold $d > 1$. The case of $d \leq 1$ is handled differently by a direct graph theoretic method, which is not relevant to our discussion here.

In this simplified version, the idea is to set $m = (1 - \varepsilon)n$ and start by revealing the top left $m \times m$ sub-matrix⁴ $A_m = A_{m,m}$ of A_n . We will then “build up” the matrix A_n by alternately adding rows and columns until we fill out all of A_n :

$$A_{m,m} \rightarrow A_{m,m+1} \rightarrow A_{m+1,m+1} \rightarrow \cdots \rightarrow A_{n,n}.$$

At each step we will upper bound a (sliding) window of δn singular values. Precisely, we define the *window of A_t at height r* to be

$$W_{t,r}(z) = -\frac{1}{n} \sum_{j=r}^{r+\delta n} \log(\sigma_{t-j}(A_t - zI_t)),$$

and look to maintain an upper bound of $W_{t,r}(z) = o_{\delta \rightarrow 0}(1)$, as t increases and r decreases.

We initialize our process by controlling the window of A_m at height $r_0 = \varepsilon^4 m$

$$W_{m,r_0}(z) = -\frac{1}{n} \sum_{j=r_0}^{r_0+\delta n} \log(\sigma_{m-j}(A_m - zI_m)).$$

³In fact, both Tao and Vu [30] as well as Rudelson and Tikhomirov [25] obtain better estimates on the least singular value than required, but this is not relevant to our discussion here.

⁴Here we let $A_{s,t}$ denote the submatrix of A_n defined by $(A_{ij})_{i \in [s], j \in [t]}$. We define $A_m = A_{m,m}$.

To deal with this initial window we prove that, for all $z \in \mathbb{C}$ outside a set of measure 0, there is a $\tau(z) > 0$ so that

$$(6) \quad \mathbb{P}\left(\sigma_{(1-\varepsilon^4)m}(A_m - zI) > \tau(z)\right) = 1 - o(1).$$

In fact, the core of this step is simply an abstract argument about eigenvalues of general matrices. For the remainder of the proof, we restrict our attention to the values of z which satisfy (6), which is enough for us: we don't mind losing control on a set of zero measure in z . Thus, by applying (6), we have the upper bound of

$$W_{m,r_0}(z) \leq \delta \varepsilon^4 (\log \tau^{-1}(z))$$

on our initial window, with high probability.

We now describe how our process evolves. Let us index time in our process by $t \in [m, n]$, so that at time t we have exposed A_t . When we progress from $t \rightarrow t + 1$ we add a column to A_t and then row, so that we have filled out A_{t+1} . We now shift the height of our window downward by one at time t , that is $r \rightarrow r - 1$, if after our column and row additions we have

$$(7) \quad W_{t+1,r-1}(z) \leq W_{t,r}(z) + \delta_r,$$

where $\delta_r > 0$ is the amount that we are willing to lose when we take on the r th smallest singular value. The key property of the δ_r is simply that $\sum_{r=0}^{r_0} \delta_r = o(1)$.

We now need to consider the probability that the good event (7) occurs, and that we actually make downward progress in our process. Here the intuition is that the addition of a row or column to A_t (or $A_{t,t+1}$) "pushes", monotonically, all of the singular values of A_t *upward*. We shall see that *often* the two singular values that are just below our current window are pushed hard enough for us to "capture", by moving our window downward, without increasing our logarithmic sum by too much. Since we also create a *new* singular value, as the dimension increases by 1, we move $r \rightarrow r - 1$ whenever (7) occurs.

In particular, we shall show that (7) occurs with decent probability: if d is large then we have

$$(8) \quad \mathbb{P}(W_{t+1,r-1}(z) \leq W_{t,r}(z) + \delta_r) = 1 - o_{d \rightarrow \infty}(1),$$

subject to A_t (and then $A_{t,t+1}$) satisfying several quasi-randomness conditions. Here the probability in (8) is only over the new row and column addition, which allows us to naturally think of the evolution the height as a random process in t . We remark that proving (8) is one of the main technical challenges of this paper. We briefly discuss the ingredients that go into proving this in Section 2.4 below.

To see that (8) ensures that we make sufficient downward progress, let X_t be the height of our window at time t . The trajectory of X_t is defined by

$$X_{t+1} \leq X_{t-1} + 1 \quad \text{if (8) fails; and } X_{t+1} = X_t - 1 \quad \text{if (8) holds.}$$

Since we have εn steps in our process; each step has downward drift of nearly 1; and our starting point is $X_m = \varepsilon^4 m < \varepsilon^4 n$, a standard martingale analysis reveals that we have sufficient downward drift to ensure $X_n = O(1)$, with high probability.

But here we encounter a difficulty right at the end of the process. The above analysis only guarantees that $X_n = O(1)$, with high probability, while we require that $X_n = 0$, with high probability. The key observation here is that a random iid Bernoulli matrix, with $p = d/n$ for $d > 1$, will actually have a few rows that have a *large* number of 1s. Thus we arrange that a few of these rows appear at the very end of the process, so that we obtain a stronger probability bound on (8) and therefore create a strong downward drift for X_t in the last few steps, which ultimately guarantees that $X_n = 0$, with high probability.

In particular, we arrange that we encounter, at the end of the process, $\ell = (\log n)^2$ iid rows with approximately $\sqrt{\log n}$ 1s in each. This allows us to ensure that

$$\mathbb{P}(W_{r-1}(z) \leq W_r(z) + \delta_r) \geq 1 - (\log n)^{-c},$$

for $c > 0$, in the last ℓ steps of the process.

2.3. Extending to the threshold. While the proof we have sketched above can be made to work for all sufficiently large d , the bound we obtain in (8) is not good enough to accommodate smaller d . To get down to the threshold $d > 1$, we need one further idea. Again the idea is to move a bunch of rows and columns with support size $\gg 1$ to the end of the process, where they are most useful. Here, however, our technique is slightly different than that described above and based on a similar manoeuvre found in [10, 13].

This idea is best explained if we think of our matrix as a directed graph on $[n]$ where (i, j) is an edge if the ij th entry is 1. We start by setting aside the first $(1 - \varepsilon)n$ vertices (which corresponds to the top left principal submatrix). We now expose the *degrees* (both in and out) of the vertices⁵ $(1 - \varepsilon)n, \dots, n - \ell$ within $[n - \ell]$ and define the *value* of a vertex $j \in \{(1 - \varepsilon)n, \dots, n - \ell\}$ to be

$$\text{val}(j) = \min\{\text{deg}^+(j, [n - \ell]), \text{deg}^-(j, [n - \ell])\}.$$

We then move the $\varepsilon^3 n$ vertices among $\{(1 - \varepsilon)n, \dots, n - \ell\}$ with the largest value to the end of our ordering. Note that we expect that these vertices will have value $\geq (\log 1/\varepsilon)^{1-o(1)}$.

Being careful that we maintain enough independence after these degree exposures, we can prove a variant of (8) of the form

$$(9) \quad \mathbb{P}(W_{t+1, r-1}(z) \leq W_{t, r}(z) + \delta_r) = 1 - o_{\varepsilon \rightarrow 0}(1),$$

where the little o terms now (crucially) tends to zero as $\varepsilon \rightarrow 0$. Hence we recover the same behavior we saw for d sufficiently large at (8).

We now run our process on the last $\varepsilon^3 n + \ell$ rows and columns, again setting aside $\ell = (\log n)^2$ rows with $\approx \sqrt{\log n}$ 1s to help us at the end. That is, we initialize our process at

$$t_0 = m = n - \varepsilon^3 n - \ell \quad \text{and with initial window height} \quad r_0 = \varepsilon^4 n$$

and then fill out A_n by adding the remaining $\varepsilon^3 n + \ell$ rows and columns.

To see that we have enough downward drift in our window height X_t , we note that $X_m = \varepsilon^4 m < \varepsilon^4 n$ and that in each of our $\varepsilon^3 n$ steps we have downward drift of nearly 1, from (9). Thus a standard martingale analysis again tells us that we have $X_{n-\ell} = O(1)$ with high probability. We can then ensure $X_n = 0$, with high probability, after exposing the last ℓ rows which we set aside previously.

2.4. The proof of (8), the key step. One of the key components of the paper is the proof of (8) and its relative (9). Our first step toward (8) is to prove the following inequality that quantifies how much our window is pushed after on the addition of a new column (or row) X :

$$(10) \quad \prod_{j=r-1}^{r-1+\delta n} \sigma_{t-j}(A_{t,t+1}) \geq \|PX\|_2 \cdot (\|X\|_2^2 + \sigma_{t-r-\delta n}(A_t)^2)^{-1/2} \cdot \prod_{j=r}^{t+\delta n} \sigma_{t-j}(A_t),$$

The key thing to observe here is that our bound crucially depends on the quantity $\|PX\|_2$, where P is the orthogonal projection onto the span of the singular directions which are below our window. In particular, small $\|PX\|_2$ is bad for us here, while large $\|PX\|_2$ is good.

Thus (10) reduces our task to proving that it is unlikely for X to have small projection onto the span of the smallest singular directions below our current window. While this is *not* true

⁵Here and throughout we use the notation $[n] = \{1, \dots, n\}$.

for a general matrix, we are able to prove that this is true for quasi-random A_t . More precisely we show that if A_t satisfies certain quasi-randomness conditions, we have

$$(11) \quad \mathbb{P}_X(\|PX\|_2 < \exp(-n\delta_r)) = o_{\varepsilon \rightarrow 0}(1).$$

While proving “unstructured” theorems of this type can be quite challenging in general, here we can get away with a very weak notion of quasi-randomness, based on a simple variant of a graph expansion property: this is our so called “unique neighbourhood expansion” property. From this quasi-randomness condition, we can deduce that all vectors that are close to the kernel of A are reasonably “flat”. This then allows us to deduce that the span of the singular directions is sufficiently quasi-random to imply (11).

2.5. Organization of the paper. In Section 3 we prove Theorem 1.1 when $d \leq 1$ using graph theoretic techniques. In Section 4 we define the extraction procedure for high degree vertices and construct the natural filtration by which to run the random walk. In this section we also define various notation regarding the random walk which will be used throughout the paper. In Section 5 we prove the coupling estimate between a $\text{Ber}(d/n)$ matrix and a matrix where the final set of rows and columns have modified entries to ensure high degree at the very end of the process.

In Section 6, we prove that various digraphs in the walk exhibit unique neighborhood expansion. In Section 7, we transfer this expansion into unstructuredness of near minimal singular vectors. In Section 8, we prove that this implies that the image of the new row or column added is unlikely to lie near the subspace spanned by near minimal singular vectors. In Section 9, we prove a basic result regarding random walks with drift.

In Section 10 we prove the crucial linear algebra lemma which relates knowledge of the projection onto near minimal singular vectors into control allowing a shift of the window of singular values downward. In Section 11 we prove convergence of the singular values of shifted random matrices. In Section 12 we prove that the shifted singular value measures have no atom at 0 for Lebesgue almost all z . Finally, in Section 13 we run the random walk in order to complete the proof by combining various ingredients proven earlier.

3. HANDLING $d \leq 1$

We prove Theorem 1.1 in the case when $d \leq 1$. The proof is purely graph theoretic. We view A_n as the $n \times n$ adjacency matrix of a digraph (see Appendix A.1). Let $\mathbb{D}(n, p)$ be the random directed graph where each directed edge (including self loops) is included with probability p . Thus, A_n is the adjacency matrix of $G' \sim \mathbb{D}(n, d/n)$. Since $d \leq 1$, we can find a joint distribution of random digraphs such that $G' \sim \mathbb{D}(n, d/n)$, $G \sim \mathbb{D}(n, (1 + \varepsilon)/n)$, and $G' \subseteq G$ almost surely. We will consider such couplings where ε is a fixed constant such that $\varepsilon \in (0, 1)$, and ultimately consider the behavior as $\varepsilon \rightarrow 0$.

We first have the following result regarding the structure of strongly connected components of $\mathbb{D}(n, p)$ coming from Łuczak [20, Theorem 1(ii)].⁶

Theorem 3.1. *Fix $\varepsilon \in (0, 1)$. Let $G \sim \mathbb{D}(n, p)$ with $p = (1 + \varepsilon)/n$. The strongly connected components of G consist of one component of size $\Omega_\varepsilon(n)$ and the remaining are either cycles of length at most $\log \log n$ or components of size 1.*

⁶Łuczak [20] and Karp [17] consider $\mathbb{D}(n, p)$ without loops but this leaves the structure of strongly connected components unchanged. Furthermore Łuczak considers $\mathbb{D}(n, m)$ but the desired result follows in an unchanged manner.

We will also require the following set of results due to Karp [17]. In order to simplify notation, we say that $u \rightarrow v$ (with respect to a directed graph G) if there exists a directed path from u to v (with the convention $u \rightarrow u$). We define

$$X_G(u) = \{v \in V(G) : u \rightarrow v\} \text{ and } Y_G(u) = \{v \in V(G) : v \rightarrow u\}.$$

We have the following results concerning $X_G(u)$ and $Y_G(v)$ due to Karp [17, Theorems 1, 3].

Theorem 3.2. *Fix $\varepsilon \in (0, 1)$. Let $G \sim \mathbb{D}(n, p)$ with $p = (1 + \varepsilon)/n$ and define $\theta = \theta(\varepsilon)$ to be the unique nonzero solution of $1 - x - e^{-(1+\varepsilon)x} = 0$. Then with probability $1 - n^{-\omega(1)}$, we have for all $v \in [n]$ that*

$$X_G(v) \in [0, (\log n)^2] \cup [\theta n - n^{1/2} \log n, \theta n + n^{1/2} \log n]$$

and by symmetry the same for $Y_G(v)$.

Furthermore, with probability $1 - n^{-\omega(1)}$, if $|X_G(u)| \geq \theta n/2$ and $|Y_G(v)| \geq \theta n/2$, then $u \rightarrow v$.

We now prove Theorem 1.1 in the case when $d \leq 1$. We will in fact prove that μ_d is an Dirac mass at 0.

Proof of Theorem 1.1 for $d \leq 1$. Fix $\varepsilon > 0$. One can jointly sample $G' \sim \mathbb{D}(n, d/n)$ and $G \sim \mathbb{D}(n, (1 + \varepsilon)/n)$ such that $G' \subseteq G$ almost surely. We define a vertex v to have *trivial image* in G' (and analogous for G) if for all $u \in X_{G'}(v)$, we have that the strongly connected component corresponding to u in G' has size 1 and u does not have a self-loop. We claim that all but at most $2\theta n$ vertices in G have trivial image whp, where $\theta = \theta(\varepsilon)$ is as in Theorem 3.2.

Given this, note that there can only be more vertices with trivial image in G' . Additionally, note that if A is the adjacency matrix corresponding to digraph G' then $A^n e_v = 0$ for all v with trivial image. Therefore we deduce that A^n is nonzero only on a dimension at most $2\theta n$ space. This immediately implies that A has at least $(1 - 2\theta(\varepsilon))n$ many eigenvalues equal to 0. Finally, note that $\theta(\varepsilon) \rightarrow 0$ as $\varepsilon \rightarrow 0$, so it follows that all but $o(1)$ fraction of the eigenvalues of A are 0. Thus the (random) spectral measure converges in probability to the Dirac mass at 0.

We now prove the above claim. Assume the outcomes of Theorems 3.1 and 3.2 hold, which occurs whp. Let S be the set of vertices in the unique giant strongly connected component guaranteed by Theorem 3.1. Let S_- be the set of vertices u for which $X_G(u) \supseteq S$ and let S_+ be the set of vertices u for which $Y_G(u) \supseteq S$. (Note $S \subseteq S_+ \cap S_-$.) Let $T = S \cup S_+ \cup S_-$. By Theorem 3.2, for all $v \in S_+$ we have $|Y_G(v)| \in [\theta n \pm n^{1/2} \log n]$.

In particular, applying this for $v \in S$ demonstrates that $|S_-| \in [\theta n \pm n^{1/2} \log n]$. Then applying it for $v \in S_+$ shows that $Y_G(v)$ contains at most $2n^{1/2} \log n$ vertices not in S_- . Furthermore, if $v \in V(G)$ satisfies $|Y_G(v)| \geq \theta n/2$ then the final part of Theorem 3.2 implies that $u \rightarrow v$ for all $u \in S$, i.e., $v \in S_+$. So, if $v \notin S_+$ then $|Y_G(v)| < \theta n/2$ hence in fact $|Y_G(v)| \in [0, (\log n)^2]$ by Theorem 3.2. Combining these cases, we have $|Y_G(v) \setminus S_-| \leq 2n^{1/2} \log n$ for all $v \in V(G)$.

Finally, consider a vertex $u \notin S_-$ which does not have trivial image. By Theorem 3.1, there must exist $v \in X_G(u)$ such that $v \in S$ or $v \notin S$ and v is in a cycle of length at most $\log \log n$ (where a self-loop is treated as a cycle of length 1). In the former case, we must have $u \in S_-$, which contradicts $u \notin S_-$. Thus the latter case holds. Additionally, note that $u \in Y_G(v) \setminus S_-$ holds.

Thus, any $u \notin S_-$ which does not have trivial image must be in $Y_G(v) \setminus S_-$ for some v in a cycle of length at most $\log \log n$. There are at most $n^{o(1)}$ such v whp, and by the previous paragraph each such v satisfies $|Y_G(v) \setminus T| \leq 2n^{1/2} \log n$. Thus u must be in a set of vertices of size at most $n^{1/2+o(1)}$.

Putting this together, we deduce there are at most $|S_-| + n^{1/2+o(1)} \leq 2\theta n$ many vertices $u \in V(G)$ that do not have trivial image. We are done. \square

4. DEFINITION OF THE RANDOM PROCESS

We now formally define the crucial sequence of random matrices which we will be concerned with for the remainder of the paper. Fix a constant $d \neq 1$; d will be fixed throughout the proof and constants will always be allowed to depend on d (specifically, its size and its distance from 1). Let $\ell = \lfloor (\log n)^2 \rfloor$.

Let A be an $n \times n$ matrix with independent $\{0, 1\}$ -entries such that

$$(12) \quad \mathbb{P}(A_{ij} = 1) = d/n.$$

Let B be an $n \times n$ matrix with independent $\{0, 1\}$ -entries such that

$$(13) \quad \mathbb{P}(B_{ij} = 1) = d/n \text{ if } \max(i, j) \leq n - \ell \text{ and } \mathbb{P}(B_{ij} = 1) = \sqrt{\log n}/n \text{ otherwise.}$$

Let $\varepsilon = \mathbb{P}(\text{Pois}(d) \geq \Delta)$ for an integer Δ to be chosen later. Throughout the proof we will ensure that various parameters are independent of Δ ; we will ultimately take $\Delta \rightarrow \infty$ at the end of the proof.

We will build B iteratively by adding columns and rows alternately. However the order that we add them in will be rather delicate. Further define

$$T_1 = [\lfloor n(1 - \varepsilon) \rfloor], \quad T_2 = [\lfloor n(1 - \varepsilon) \rfloor + 1, n - \ell], \quad \text{and } T_3 = [n - \ell + 1, n].$$

We now define the set of high-degree indices in T_2 . For each index j in T_2 , define the *value* of j to be

$$\text{val}(j) = \min(\text{deg}_B^+(j, [n - \ell]), \text{deg}_B^-(j, [n - \ell])),$$

where we associate $\{0, 1\}$ -matrices to digraphs as outlined in Appendix A.1. (We will use the conventions there in the remainder of the paper without further mention.)

Let H be defined as the $\lfloor \varepsilon^3 n \rfloor$ indices j in T_2 with largest $\text{val}(j)$ (breaking ties by choosing earlier indices in the integer ordering first).

We now iteratively build our matrix B such that $B_n = B$; the precise ordering of indices will depend on the parameter ε . Define $m = m(\varepsilon) = n - \ell - \lfloor \varepsilon^3 n \rfloor$. Then define a sequence of random sets S_j for $m \leq j \leq n$ and random indices v_j for $m + 1 \leq j \leq n$ iteratively as

$$\begin{aligned} S_m &= T_1 \cup (T_2 \setminus H), \\ v_{j+1} &= \text{Unif}(H \setminus S_j), & \text{for } m \leq j < n - \ell, \\ v_{j+1} &= j, & \text{for } n - \ell \leq j < n, \\ S_{j+1} &= S_j \cup \{v_{j+1}\}, & \text{for } m \leq j < n. \end{aligned}$$

Finally, define

$$B_j = B[S_j, S_j] \text{ for } m \leq j \leq n \text{ and } B_j^* = B[S_{j-1}, S_j] \text{ for } m + 1 \leq j \leq n.$$

Less formally, for the last ℓ steps of the process we build the matrix B by adding back in columns and then rows in the obvious manner. For the first $\lfloor \varepsilon^3 n \rfloor$ steps of the process though, we extract the vertices in T_2 with high in-degree and out-degree (with respect to the digraph given by restricting attention $T_1 \cup T_2$), and add them back in a random order. Given this it is natural to define $m \leq j \leq n - \ell$ as the *first epoch* and $n - \ell + 1 \leq j \leq n$ as the *second epoch*.

We state a series of elementary properties regarding the random walk and matrices B_t .

Fact 4.1. *We have $S_{n-\ell} = [n - \ell]$ deterministically. Furthermore, indices $B_{i,j}$ for $\max(i, j) \geq n - \ell + 1$ are jointly independent of B_t, B_t^* for $t \leq n - \ell$.*

Proof. The fact $S_{n-\ell} = [n - \ell]$ follows by construction of the sequence v_j . For the second part of the claim, note that the ordering of indices to form $S_{n-\ell}$ is only dependent on the entries of $B_{i,j}$ with $\max(i, j) \leq n - \ell$ (and independent randomness). Since B_t, B_t^* are submatrices of

$B_{n-\ell}$ defined in terms of randomness independent from the last ℓ rows and columns, the desired result follows from the initial definition of B . \square

We now state the state the obvious but crucial symmetry property of B .

Fact 4.2. *Conditional on the value $\sum_{1 \leq i, j \leq n-\ell} B_{ij}$, the matrix $B_{n-\ell}$ is uniform over all $(n-\ell) \times (n-\ell)$ matrices with $\{0, 1\}$ entries and exactly $\sum_{1 \leq i, j \leq n-\ell} B_{ij}$ many 1s.*

We next prove that the distribution of B_i is a mixture of degree-constrained random graphs. This property will be used to establish unique-neighborhood expansion facts.

Fact 4.3. *The set H is measurable given the degree sequence $(\mathbf{d}_{n-\ell}, \mathbf{d}'_{n-\ell})$ of $B_{n-\ell}$. Additionally, given j such that $m \leq j \leq n-\ell$, conditional on the index set S_j , and conditional on the degree sequence $(\mathbf{d}_j, \mathbf{d}'_j)$ of B_j , the random variable B_j is a uniformly random bipartite graph with degree sequence $(\mathbf{d}_j, \mathbf{d}'_j)$.*

Remark. The analogous result holds for B'_j ; this will not be required for the proof.

Proof. The first claim follows as H is defined via examining $\text{val}(j)$ for $j \in T_2$, which is measurable in terms of the degree sequence of $B_{n-\ell}$ by construction. Since $\sum_{1 \leq i, j \leq n-\ell} B_{i,j}$ is measurable in terms of the degree sequence of $B_{n-\ell}$, applying Fact 4.2 we have that $B_{n-\ell}$ is a uniformly random digraph given its degree sequence. Note that the sets S_j are determined given the degree sequence of $B_{n-\ell}$ and independent randomness. Furthermore, if we reveal the degree sequence of $B_j = B_{n-\ell}[S_j, S_j]$ then the conditional digraph B_j is independent of the “outside” digraph corresponding to edges of $B_{n-\ell}$ not fully contained in S_j . These facts allow us to deduce the second claim. \square

We now define the filtration \mathcal{F}_j under which the random walk will occur. We abuse notation and identify a σ -algebra with a collection of random variables; such a σ -algebra is thus defined as the minimal σ -algebra such that the collection of random variables is jointly measurable. Define the set of pairs of indices

$$\mathcal{R} = (S_m \times S_m) \cup (T_2 \times H) \cup (H \times T_2).$$

\mathcal{R} will correspond to the set of revealed entries of B for the initial setup of the random walk. In particular \mathcal{R} reveals all edges in the initial S_m and all edges between H and T_2

Define

$$\begin{aligned} \mathcal{F}_m &= \{B_{ij}\}_{(i,j) \in \mathcal{R}} \cup \{\deg_B^+(i, [n-\ell]), \deg_B^-(i, [n-\ell])\}_{i \in T_2}, \\ \mathcal{F}_j &= \mathcal{F}_{j-1} \cup \{S_{j-1}, B_{j-1}\}, & \text{for } m+1 \leq j < n, \\ \mathcal{F}'_j &= \mathcal{F}_j \cup \{S_j\}, & \text{for } m+1 \leq j < n. \end{aligned}$$

\mathcal{F}_m can be viewed as revealing all edges of $B_{n-\ell}$ in \mathcal{R} and the degrees of all vertices in T_2 with respect to the digraph $B_{n-\ell}$. \mathcal{F}_j is obtained from \mathcal{F}_{j-1} by revealing the vertex v_{j-1} and its edges to all vertices in S_{j-1} . \mathcal{F}'_j is obtained from \mathcal{F}_j by revealing the identity of v_j . (So, \mathcal{F}_{j+1} compared to \mathcal{F}'_j is only additionally revealing the in- and out-neighborhood of v_j within B_j .)

We now state a series of claims regarding the σ -algebra \mathcal{F}_j which can be easily seen by chasing definitions.

Fact 4.4. *We have the following:*

- \mathcal{F}_j for $m \leq j \leq n$ forms a filtration;
- H is measurable with respect to \mathcal{F}_m ;
- $\deg_B^+(i, T_1)$ and $\deg_B^-(i, T_1)$ for $i \in T_2$ are measurable with respect to \mathcal{F}_m ;
- $(B_j)_{k,\ell}$ are \mathcal{F}'_j -measurable except if $v_j \in \{k, \ell\}$ and $\{k, \ell\} \cap T_1 \neq \emptyset$;

- $\deg_{B_j}^+(v_j)$ and $\deg_{B_j}^-(v_j)$ are \mathcal{F}'_j -measurable.

The last bullet point follows since \mathcal{F}_m reveals \mathcal{R} , which includes all pairs in $H \times H$. Now, the crucial property of the sequence of σ -algebras \mathcal{F}_j is that the “remaining” randomness of B_j given \mathcal{F}_j and S_j is particularly simple.

Fact 4.5. *Given \mathcal{F}'_j we have that $\{(B_j)_{v_j,i}\}_{i \in T_1}$ and $\{(B_j)_{i,v_j}\}_{i \in T_1}$ are independent uniformly random $\{0,1\}$ -vectors conditional on the sums $\sum_{i \in T_1} (B_j)_{i,v_j}$ and $\sum_{i \in T_1} (B_j)_{v_j,i}$, respectively. (These sums are deterministic given \mathcal{F}'_j , by the last bullet point of Fact 4.4.)*

5. COUPLING ESTIMATES

We first require the following coupling estimate between A and B .

Lemma 5.1. *Let A and B be as in (12) and (13) and let P_σ be a uniformly random $n \times n$ permutation matrix. For n sufficiently large,*

$$\text{TV}(A, P_\sigma^T B P_\sigma) \leq n^{-1/4}.$$

Lemma 5.1 follows immediately by iterating the following result, which states that the total variation distance is small if we replace a random $\text{Ber}(d/n)$ row and column with a random $\text{Ber}(\sqrt{\log n}/n)$ row and column. We remark that beyond implying Lemma 5.1, Lemma 5.2 in fact shows that one can somewhat freely pass permutation-symmetric events that hold whp between an independent model with probability $p = d/n$ and an independent model where a small number of rows or columns have an altered probability; we will briefly remark when we are doing this.

Lemma 5.2. *Suppose that $d \in [(\log n)^{-1/3}, (\log n)^{1/3}]$ and $\tau \in [(\log n)^{1/2}/(2n), 2(\log n)^{1/2}/n]$. Define probability distributions $\mathcal{P}_1, \mathcal{P}_2$ on $n \times n$ $\{0,1\}$ -matrices M :*

- \mathcal{P}_1 : Each entry M_{ij} is 1 with probability d/n independently at random.
- \mathcal{P}_2 : Let σ be a uniformly random element of $[n]$. If $\sigma \in \{i, j\}$, then $M_{ij} = 1$ with probability $\tau \in [(\log n)^{1/2}/(2n), 2(\log n)^{1/2}/n]$; else $M_{ij} = 1$ with probability d/n .

We have

$$\text{TV}(\mathcal{P}_1, \mathcal{P}_2) \leq n^{-1/3+o(1)}.$$

Proof. Let $L = (\log n)(\log \log n)^{-1/2}$ and note that

$$\mathbb{P}_{\mathcal{P}_1} \left(\max_{i \in [n]} \sum_{j=1}^n M_{ij} \geq L \right) \leq n \binom{n}{L} \left(\frac{d}{n} \right)^L \leq n \cdot \left(\frac{en}{L} \right)^L \left(\frac{d}{n} \right)^L \leq n^{-\omega(1)},$$

and similar for columns. For a matrix M , let

$$S_\ell(M) = \# \left\{ k : \sum_{i=1}^n M_{ik} + \sum_{j=1}^n M_{kj} - M_{kk} = \ell \right\}$$

and note for $\ell \leq 2L$ that

$$\mathbb{E}_{\mathcal{P}_1} S_\ell(M) = n \binom{d}{n}^\ell \binom{2n-1}{\ell} \left(1 - \frac{d}{n} \right)^{2n-1-\ell} = (1 \pm n^{-1+o(1)}) \frac{n(2d)^\ell e^{-2d}}{\ell!}$$

and

$$\begin{aligned} \mathbb{E}_{\mathcal{P}_1} S_\ell(M)^2 &\leq \mathbb{E}_{\mathcal{P}_1} S_\ell(M) + (1 \pm n^{-1+o(1)}) n^2 \left(\frac{(2d)^\ell}{\ell!} e^{-2d} \right)^2 \\ &\leq (1 \pm n^{-1+o(1)}) (\mathbb{E}_{\mathcal{P}_1} S_\ell(M) + (\mathbb{E}_{\mathcal{P}_1} S_\ell(M))^2). \end{aligned}$$

Thus by Chebyshev's inequality, we have for all $0 \leq \ell \leq 2L$ that

$$|S_\ell(M) - \mathbb{E}_{\mathcal{P}_1} S_\ell(M)| \leq n^{1/6+o(1)} \sqrt{\mathbb{E}_{\mathcal{P}_1} S_\ell(M)} + n^{-1/3+o(1)} \mathbb{E}_{\mathcal{P}_1} S_\ell(M)$$

with probability at least $1 - n^{-1/3}$. Let \mathcal{G} denote the set of matrices $M \in \{0, 1\}^{n \times n}$ such that

$$|S_\ell(M) - \mathbb{E}_{\mathcal{P}_1} S_\ell(M)| \leq n^{1/6+o(1)} \sqrt{\mathbb{E}_{\mathcal{P}_1} S_\ell} + n^{-1/3+o(1)} \mathbb{E}_{\mathcal{P}_1} S_\ell$$

for $1 \leq \ell \leq 2L$ and $S_\ell(M) = 0$ for $\ell > 2L$ hold simultaneously. Therefore we find

$$\begin{aligned} \text{TV}(\mathcal{P}_1, \mathcal{P}_2) &= \sum_{M' \in \{0,1\}^{n \times n}} (\mathbb{P}_{M \sim \mathcal{P}_1}(M = M') - \mathbb{P}_{M \sim \mathcal{P}_2}(M = M')) \mathbb{1}_{\mathbb{P}_{M \sim \mathcal{P}_1}(M=M') \geq \mathbb{P}_{M \sim \mathcal{P}_2}(M=M')} \\ &\leq \sum_{\substack{M' \in \{0,1\}^{n \times n} \\ M' \in \mathcal{G}}} \left| \mathbb{P}_{M \sim \mathcal{P}_1}(M = M') - \mathbb{P}_{M \sim \mathcal{P}_2}(M = M') \right| + \mathbb{P}_{M \sim \mathcal{P}_1}(A \notin \mathcal{G}) \\ &\leq \sum_{\substack{M' \in \{0,1\}^{n \times n} \\ M' \in \mathcal{G}}} \mathbb{P}_{M \sim \mathcal{P}_1}(M = M') \cdot \left| 1 - \frac{\mathbb{P}_{M \sim \mathcal{P}_2}(M = M')}{\mathbb{P}_{M \sim \mathcal{P}_1}(M = M')} \right| + n^{-1/3}. \end{aligned}$$

However, for all $M' \in \mathcal{G}$ we have

$$\begin{aligned} \left| 1 - \frac{\mathbb{P}_{M \sim \mathcal{P}_2}(M = M')}{\mathbb{P}_{M \sim \mathcal{P}_1}(M = M')} \right| &= \left| 1 - \frac{1}{n} \sum_{0 \leq \ell \leq 2L} S_\ell(M') \left(\frac{\sqrt{\log n}}{d} \right)^\ell \left(1 - \frac{\sqrt{\log n}}{n} \right)^{2n-1-\ell} \left(1 - \frac{d}{n} \right)^{-(2n-1-\ell)} \right| \\ &\leq \left| 1 - \frac{1}{n} \sum_{0 \leq \ell \leq 2L} \frac{n(2d)^\ell e^{-2d}}{\ell!} \left(\frac{\sqrt{\log n}}{d} \right)^\ell e^{-2\sqrt{\log n}+2d} \right| + n^{-1+o(1)} \\ &\quad + \frac{1}{n} \sum_{0 \leq \ell \leq 2L} \left(n^{1/6+o(1)} \left(\frac{n(2d)^\ell e^{-2d}}{\ell!} \right)^{1/2} + n^{-1/3+o(1)} \left(\frac{n(2d)^\ell e^{-2d}}{\ell!} \right) \right) \left(\frac{\sqrt{\log n}}{d} \right)^\ell e^{-2\sqrt{\log n}+2d} \\ &\leq \left| 1 - \frac{1}{n} \sum_{0 \leq \ell \leq 2L} \frac{n(2d)^\ell e^{-2d}}{\ell!} \left(\frac{\sqrt{\log n}}{d} \right)^\ell e^{-2\sqrt{\log n}+2d} \right| + n^{-1/3+o(1)} \leq n^{-1/3+o(1)}. \end{aligned}$$

Therefore

$$\text{TV}(\mathcal{P}_1, \mathcal{P}_2) \leq \sum_{\substack{M' \in \{0,1\}^{n \times n} \\ M' \in \mathcal{G}}} \mathbb{P}_{M \sim \mathcal{P}_1}(M = M') \cdot \left| 1 - \frac{\mathbb{P}_{M \sim \mathcal{P}_2}(M = M')}{\mathbb{P}_{M \sim \mathcal{P}_1}(M = M')} \right| + n^{-1/3+o(1)} \leq n^{-1/3+o(1)},$$

and we are done. \square

6. UNIQUE NEIGHBOURHOOD EXPANSIONS

We now define the set of unique neighbors of a set S for a $\{0, 1\}$ -matrix. Let M be an $m \times \ell$ matrix and let $S \subseteq [\ell]$ be a set of columns of M . We define $U(S) \subseteq [m]$ to be a subset of rows in two stages. We first define

$$U(S) \setminus S = \{i \in [m] \setminus S : B_{ij} = 1 \text{ for a unique } j \in S\}.$$

We then define

$$U(S) \cap S = \{i \in [m] \cap S : B_{ij} = 0 \text{ for all } j \in S\}.$$

We fix $\alpha(x) = (\log(n/x))^{-2}$ (here n is taken to be the same as the global parameter n defining A). The key output of this section will (essentially) be proving with high probability that we may assume that $|U(S)| \geq \alpha(|S|)|S|$ for S of the relevant size in the process.

6.1. Calculations in the independent model. We now bound the number of small sets which do not have many unique neighbors; the analysis here is nearly identical to that in our companion paper [27, Lemma 4.2].

Lemma 6.1. *Fix $\delta > 0$, consider integers m and $\ell \in \{m, m+1\}$, and choose $pm \in [1 + \delta, \delta^{-1}]$. Suppose M is an $m \times \ell$ matrix with iid $\text{Ber}(p)$ entries, and assume $m/n \in [1/2, 2]$. There exist constants $c > 0$, $C > 0$ depending only on δ such that the following holds.*

For all $k \in [0, cn]$, we have

$$\mathbb{E}_M \left| \left\{ S \in \binom{[\ell]}{k} : |U(S)| < \alpha(k)k \right\} \right| \leq C e^{-ck}.$$

Remark. The assumption that $pm \geq 1 + \delta$ is used in a crucial manner although the proof may be adjusted to handle $pm \leq 1 - \delta$.

Proof. For the sake of simplicity we will consider the case when $\ell = m + 1$; the other case is strictly simpler. Fix a set $S \subseteq [\ell]$ of columns of size k . Note that

$$|U(S)| = \sum_{i \in [m]} \mathbb{1}(i \in U(S)),$$

where the sum is over the rows and therefore a sum of independent random variables. Note that for $i \in [\ell]$ we have

$$(14) \quad \mathbb{P}(i \in U(S)) = (1-p)^k =: q_1 \text{ if } i \in S \text{ and } \mathbb{P}(i \in U(S)) = (1-p)^{k-1}pk =: q_2 \text{ if } i \notin S.$$

Let $k' = |S \cap [m]| \in \{k-1, k\}$ and set $T = \alpha(k)k$. Now $|U(S) \cap S|$ is distributed as binomial random variable $B(k', q_1)$ and $|U(S) \setminus S|$ is distributed as $B(m - k', q_2)$. Note by Bernoulli's inequality that $1 - q_1 \leq pk$ and $q_2 \geq (1-p)^k pk \geq (1-pk)pk$. Take η to be a sufficiently small constant with respect to δ to be chosen later. By taking c sufficiently small in terms of η , we have $pk \leq \eta$ and $k \leq \eta m$.

Therefore we have

$$\begin{aligned} \mathbb{E} \left| \left\{ S \in \binom{[\ell]}{k} : |U(S)| < T \right\} \right| &\leq \mathbb{E} \left| \left\{ S \in \binom{[\ell]}{k} : |U(S) \cap S| < T, |U(S) \setminus S| < T \right\} \right| \\ &\leq \sum_{\substack{i, j < T \\ k' \in \{k-1, k\}}} \binom{m}{k'} \mathbb{P}(B(k', q_1) = i) \cdot \mathbb{P}(B(m - k', q_2) = j) \\ &\leq \sum_{\substack{i, j < T \\ k' \in \{k-1, k\}}} \binom{m}{k'} \binom{k'}{i} (1 - q_1)^{k'-i} \cdot \binom{m - k'}{j} (1 - q_2)^{m - k' - j} \\ &\leq (T + 1)^2 \binom{m}{\lfloor T \rfloor}^2 \sum_{k' \in \{k-1, k\}} \binom{m}{k'} (pk)^{k'-T} \cdot (1 - pk + (pk)^2)^{m - k' - T} \\ &\leq 2(T + 1)^2 \binom{m}{\lfloor T \rfloor}^2 (pk)^{-2T} (emp)^k \cdot (1 - pk + (pk)^2)^{m - k}. \end{aligned}$$

Using $\binom{a}{b} \leq \left(\frac{ea}{b}\right)^b$ and $1 - x \leq e^{-x}$, we have

$$\begin{aligned} 2(T + 1)^2 \binom{m}{\lfloor T \rfloor}^2 (pk)^{-2T} (emp)^k \cdot (1 - pk + (pk)^2)^{m - k} \\ \leq \exp(O(T \log(m/T))) \cdot (emp)^k \cdot (1 - pk(1 - \eta))^{m(1 - \eta)} \\ \leq \exp(O(T \log(m/T))) \cdot (emp)^k e^{-pkm(1 - \eta)^2} \end{aligned}$$

$$= \exp(O(T \log(m/T))) \cdot (empe^{-pm(1-\eta)^2})^k \leq Ce^{-ck}.$$

The final line follows since if $\delta \gg \eta \gg c$ then for $x \geq 1 + \delta$, we have $ex \leq e^{x(1-\eta)^2 - 2cx}$. \square

We will also require the following exceptional case which is used to handle steps near the end of the process.

Lemma 6.2. *Fix $\delta > 0$ and consider integers m, ℓ such that $\ell = m + 1$. There exists $C > 0$ depending only on δ such that the following holds. Let $pm \in [1 + \delta, \delta^{-1}]$, $\tau m \in [\sqrt{\log m}/2, 2\sqrt{\log m}]$, and let M have independent $\{0, 1\}$ -entries such that $M_{ij} \sim \text{Ber}(p)$ for $j \in [m]$ and $M_{i\ell} \sim \text{Ber}(\tau)$. For $k \in [0, (\log m)^{1/3}]$, we have*

$$\mathbb{E}_M \left| \left\{ S \in \binom{[\ell]}{k+1} : \ell \in S \text{ and } |U(S)| = 0 \right\} \right| \leq C \exp(-(\log m)^{1/2}/16).$$

Proof. As in Lemma 6.1, we have

$$\begin{aligned} \mathbb{P}(i \in U(S)) &\geq (1-p)^k(1-\tau) \geq 1-pk-\tau && \text{if } i \in S, \\ \mathbb{P}(i \in U(S)) &\geq \mathbb{P}(M_{i\ell} = 1) \cdot (1-p)^k \geq \sqrt{\log m}/(4m) && \text{if } i \notin S. \end{aligned}$$

Therefore

$$\begin{aligned} \mathbb{E}_M \left| \left\{ S \in \binom{[\ell]}{k+1} : \ell \in S, |U(S)| = 0 \right\} \right| &\leq \binom{m}{k} \cdot (pk + \tau)^k (1 - \sqrt{\log m}/(4m))^{m-k} \\ &\leq e^k (m/k)^k \cdot \left(\frac{3\sqrt{\log m}}{m} \right)^k \exp(-\sqrt{\log m}/8) \leq \exp(-(\log m)^{1/2}/16), \end{aligned}$$

where we have used that m is sufficiently large with respect to δ^{-1} hence $pk \leq \sqrt{\log m}$. \square

6.2. Calculations in the configuration model. The analysis in Lemmas 6.1 and 6.2 will be sufficient to analyze the unstructured of vectors arising from the second epoch. For the first epoch, we will require a more delicate analysis based on the configuration model. We first state the precise regularity conditions on the degree sequence that will be required.

Definition 6.3. *Consider degree sequences \mathbf{d}, \mathbf{d}' both of length m . We say the degree sequence $(\mathbf{d}, \mathbf{d}')$ is (d, μ, C) -regular if:*

- $m/n \in [1 - \mu, 1 + \mu]$;
- $\sum_{i \in S} (d_i + d'_i) \leq C(d + \log(m/|S|))|S|$ for all $S \subseteq [m]$;
- $\sum_{i=1}^m d_i = \sum_{i=1}^m d'_i = (1 \pm \mu)dm$;
- $\sum_{i=1}^m d^{-d_i} d'_i = (1 \pm \mu)ed \exp(-d)m$.

We will use \mathbf{d} to denote the degree sequence of the columns and \mathbf{d}' for the rows.

Lemma 6.4. *Fix $\delta > 0$ and $C > 0$. There exists $c = c(\delta, C) > 0$ such that the following holds. Let $d \in [1 + \delta, \delta^{-1}]$ and $(\mathbf{d}, \mathbf{d}')$ be a degree sequence which is (d, c, C) -regular.*

Let M be sampled as a uniformly random bipartite graph with degree sequence $(\mathbf{d}, \mathbf{d}')$ and identify M as an bipartite adjacency matrix.

For $k \in [0, cm]$, let \mathcal{Y}_k be the collection of sets S of size k satisfying

- $|\sum_{i \in S} d_i - k| \leq k/\sqrt{\log(m/k)}$;
- $|U(S)| \leq \alpha(k)k$.

Then

$$\mathbb{E}|\mathcal{Y}_k| \leq c^{-1} \exp(-ck).$$

Proof. To perform our calculations, we will operate in the configuration model. Using the second and third bullet points of Definition 6.3 and by Lemma A.6, we have that the associated configuration model is simple with probability $\Omega_{C,d}(1)$ and therefore it suffices to prove the result where the graph associated to M is sampled from the configuration model.

We first bound the probability of the event $S \in \mathcal{Y}_k$ for a specific subset $S \subseteq [m]$. Define

$$\begin{aligned} V &= \{i: d'_i \geq (\log(m/k))^3\}, \\ U_1 &= S \cap ([m] \setminus U(S)) \cap ([m] \setminus V), \\ U_2 &= ([m] \setminus S) \cap ([m] \setminus U(S)) \cap ([m] \setminus V); \end{aligned}$$

by the second item of Definition 6.3 we have that $|V| \leq \alpha(k)k$.

In order to compute the probability that $S \in \mathcal{Y}_k$, we seek to understand the probability that $j \in U_1$ has no neighbors in S while for $j \in U_2$ we need to understand the probability that it has exactly one neighbor in S . Then we will take a union bound over possible revelations of U_1, U_2 and study the chance that neither of these happen over all relevant indices.

It suffices to understand the number of stubs attached to each vertex on the right which connect to S . This distribution is given precisely by choosing each stub on the right to connect to S independently with probability $(\sum_{i \in S} d_i) / (\sum_{i \in [m]} d_i)$ and conditioning on exactly $\sum_{i \in S} d_i$ stubs being chosen. Note that the probability that the associated binomial distribution has value exactly $\sum_{i \in S} d_i$ is $\Omega(1/k)$.

Let $q = (\sum_{i \in S} d_i) / (\sum_{i \in [m]} d_i)$ and $c' = (\log(1/c))^{-1/4}$. By the first item in the lemma assumptions and the third item of Definition 6.3, we see that $q \in (1 \pm c')k / (dm)$. For given possible values of U_1, U_2 , the probability that $S \in \mathcal{Y}_k$ is bounded by

$$\begin{aligned} O_{C,d}(k) \cdot \prod_{i \in U_1} (qd'_i) \prod_{i \in U_2} (1 - qd'_i(1 - q)^{d'_i - 1}) &\lesssim_{C,d} (1 + 2c')^k \prod_{i \in U_1} \frac{kd'_i}{dm} \prod_{i \in U_2} (1 - qd'_i + (qd'_i)^2) \\ &\lesssim_{C,d} e^{2c'k} \prod_{i \in U_1} \frac{kd'_i}{dm} \prod_{i \in U_2} \left(1 - (1 - 2c') \frac{kd'_i}{dm}\right) \\ &\lesssim_{C,d} e^{2c'k} \prod_{i \in U_1} \frac{kd'_i}{dm} \cdot \exp\left(- (1 - 2c') \sum_{i \in U_2} \frac{kd'_i}{dm}\right). \end{aligned}$$

Next note that as c is sufficiently small with respect to C and d , we have

$$\sum_{i \in [m] \setminus U_2} \frac{kd'_i}{dm} \leq \frac{2k}{dm} \cdot C(d + \log(m/k))k \leq c^{1/2}k.$$

Note that we are using that $[m] \setminus U_2 \leq |S| + |V| + |U(S)| \leq 2k$ by assumption.

Now fix a constant η sufficiently small with respect to δ , and suppose c is chosen sufficiently small with respect to η . Note that $|S \setminus U_1| \leq |V| + |U(S)| \leq 2\alpha(k)k$. It follows that

$$\begin{aligned} e^{2c'k} \prod_{i \in U_1} \frac{kd'_i}{dm} \cdot \exp\left(- (1 - 2c') \sum_{i \in U_2} \frac{kd'_i}{dm}\right) &\leq e^{3c'k} \prod_{i \in U_1} \frac{k(d'_i + \eta)}{dm} \cdot \exp\left(- (1 - 2c') \sum_{i \in [m]} \frac{kd'_i}{dm}\right) \\ &\leq e^{6c'k} \prod_{i \in S} \frac{k(d'_i + \eta)}{dm} \cdot \prod_{i \in S \setminus U_1} \frac{dm}{k\eta} \cdot \exp(-k) \\ &\leq e^{6c'k - k} \prod_{i \in S} \frac{k(d'_i + \eta)}{dm} \cdot \left(\frac{dm}{k\eta}\right)^{2\alpha(k)k} \leq e^{7c'k - k} \prod_{i \in S} \frac{k(d'_i + \eta)}{dm} \end{aligned}$$

Since $|S \setminus U_1|, |(m \setminus U(S)) \setminus U_2| \leq 2\alpha(k)k$ by $S \in \mathcal{Y}_k$, we have that there are at most

$$\sum_{j \leq 2\alpha(k)k} \binom{k}{j} \cdot \sum_{j \leq 2\alpha(k)k} \binom{m}{j} \leq \left(\frac{em}{k\alpha(k)} \right)^{5\alpha(k)k} \leq \exp(c'k)$$

choices for U_1, U_2 (given S). Therefore by the union bound

$$\mathbb{P}(S \in \mathcal{Y}_k) \lesssim_{C,d} e^{8c'k-k} \prod_{i \in S} \frac{k(d'_i + \eta)}{dm}.$$

We now use the Rankin trick in order to bound the sum over all S . Namely, we use the inequality

$$\sum_{S \in \binom{[m]}{k}} \prod_{i \in S} y_i \leq \frac{1}{k!} \left(\sum_{i=1}^m y_i \right)^k$$

with the choices $y_i = d^{-d_i} k(d'_i + \eta)/(dm)$. From this we deduce

$$d^{-k-(c')^2k} \mathbb{E}|\mathcal{Y}_k| \lesssim_{C,d} e^{8c'k-k} \cdot \frac{1}{k!} \left(\sum_{i=1}^m d^{-d_i} \frac{k(d'_i + \eta)}{dm} \right)^k;$$

note that if $S \in \mathcal{Y}_k$ is counted then $\sum_{i \in S} d_i \leq k + (c')^2k$. By the third item of Definition 6.3 recall $\sum_{i=1}^m d^{-d_i} d'_i = (1 \pm c)ed \exp(-d)m$. Additionally, Stirling's formula shows $k! \lesssim k(k/e)^k$. Recalling $k \leq cm$ and that c' is small with respect to d , we find

$$\mathbb{E}Y_k \lesssim_{C,d} e^{9c'k} ((1+c)ed \exp(-d) + \eta)^k.$$

Now since $c, c' \ll \eta \ll \delta$ we have $(1+c)ed \exp(-d) + \eta \leq \exp(-9c' - c)$ for all $d \geq 1 + \delta$ (since $ex \exp(-x) < 1$ for $x > 1$, similar to the conclusion of the proof of Lemma 6.1). This concludes the proof. \square

6.3. Regularity of degree sequences. We now check various properties of the degree sequence.

Definition 6.5. Let Δ, ε be as in Section 4 and

$$\gamma = \varepsilon \cdot \frac{\mathbb{E}_{X,Y \sim \text{Pois}(d)} [X \mathbb{1}_{\min(X,Y) \geq \Delta}]}{d}.$$

For $j, k \geq 0$, let

$$\rho_{j,k}(\Delta) = (1 - \varepsilon^3)^{-1} \left(\sum_{a \geq j} \sum_{a' \geq k} \binom{\ell}{j} \binom{\ell'}{k} \gamma^{a+a'-j-k} (1 - \gamma)^{j+k} ((1 - \varepsilon) + \varepsilon \mathbb{1}_{\min(a,a') < \Delta}) \frac{d^a e^{-d}}{a!} \frac{d^{a'} e^{-d}}{a'!} \right).$$

Lemma 6.6. Fix $\Delta \geq 1$ and suppose $\varepsilon = \mathbb{P}(\text{Pois}(d) \geq \Delta)$. With probability at least $1 - n^{-\omega(1)}$, we have:

- For all $x, y \geq 0$,

$$\#\{i \in [n] : (\deg_A^+(i), \deg_A^-(i)) = (x, y)\} = \frac{d^x e^{-d}}{x!} \frac{d^y e^{-d}}{y!} n + O(n^{1/2+o(1)}).$$

- For all $x, y \geq 0$,

$$\#\{i \in V(B_m) : (\deg_{B_m}^+(i), \deg_{B_m}^-(i)) = (x, y)\} = \rho_{x,y}(\Delta) m + O(n^{1/2+o(1)}).$$

Proof. The first item follows from straightforward concentration arguments. We use permutation concentration. We can condition on the number of edges in A , which is $dn + O(\sqrt{n} \log n)$ with probability $1 - n^{-\omega(1)}$. Then the probability a single vertex has degree (x, y) is $(d^x e^{-d}/x!)(d^y e^{-d}/y!) + O(n^{-1/2+o(1)})$, so the expectation matches our prediction. Finally, we can use Lemma A.3 to obtain the desired result, noting that we can view this conditioned

model as a uniformly random injection of approximately dn edges into n^2 total possibilities, and noting that changing an edge changes our statistic by at most 4.

In fact, by the same argument we may deduce that the first $n - \ell - \lfloor \varepsilon n \rfloor$ vertices and the next $\lfloor \varepsilon n \rfloor$ vertices have the same in/out-degree distribution in this sense within the random model $B_{n-\ell}$.

We now deduce the second item from this fact. After revealing the degrees within $B_{n-\ell}$, we know the digraph is uniform over digraphs with this degree sequence. We thus use the configuration model, which we can easily check succeeds with positive probability using Lemma A.6. We now work within the digraph configuration model to prove the desired result.

For every vertex in $\{n - \ell - \lfloor \varepsilon n \rfloor + 1, \dots, n - \ell\}$ with minimum of in-degree and out-degree at least Δ , mark it red. Then mark its incoming and outgoing edges red. Let the rest of the edges and vertices of $B_{n-\ell}$ be green. We care about the green degree distribution among the vertices left after deleting the red vertices: with probability $1 - n^{-\omega(1)}$ this recovers B_m except that we may delete $O(n^{1/2+o(1)})$ more or fewer vertices than intended. With probability $1 - n^{-\omega(1)}$ the maximum degree of $B_{n-\ell}$ is at most $\log n$, so such deletions affect the degree statistics by an error of at most $O(n^{1/2+o(1)})$. So, let us focus on this “red-deletion” model, which has adjacency matrix M .

We claim that the number of vertices that go from having degrees (a, a') in $B_{n-\ell}$ to having degrees (x, y) in this final green digraph M is

$$\binom{a}{x} \gamma^{a-x} (1-\gamma)^x \binom{a'}{y} \gamma^{a'-y} (1-\gamma)^y ((1-\varepsilon) + \varepsilon \mathbb{1}_{\min(a,a') < \Delta}) \frac{d^a e^{-d}}{a!} \frac{d^{a'} e^{-d}}{a'!} n + O(n^{1/2+o(1)})$$

with probability $1 - n^{-\omega(1)}$.

Indeed, note that a $\gamma + O(n^{-1/2+o(1)})$ fraction of out-stubs are red (call this fraction γ'), and the same for in-stubs, with appropriately high probability. To study the configuration model, we care about a uniformly random matching of out-stubs and in-stubs, and we specifically consider the number of edges at each green vertex that are not partially formed from a red stub.

The joint distribution of green in/out-degree statistics computed over green vertices can easily be compared to the model where among the green vertices, each stub is retained with probability $1 - \gamma'$ independently. We start with roughly $(1 - \varepsilon)n$ vertices which are definitely green (which contributes roughly fraction $(d^a e^{-d}/a!)(d^{a'} e^{-d}/a'!)$ to the count of green vertices of total in/out-degree (a, a')) and roughly εn vertices, of which the same fraction is contributed as long as $\mathbb{1}_{\min(a,a') < \Delta} = 1$. Then we retain the resulting stubs as green independently with probability $1 - \gamma'$, which leads to the above estimate for the number of green vertices with green in/out-degree (x, y) with probability $1 - n^{-\omega(1)}$. This finishes the proof of the claim.

Now, summing over in/out-degrees (a, a') of magnitude at most $\log n$ (which the maximum degree is bounded by with probability $1 - n^{-\omega(1)}$), we obtain the desired result, recalling that $(1 - \varepsilon^3)^{-1} m \approx n$. \square

We will need the following claim, which essentially captures that the neighborhoods of the vertices in $H = V(B_{n-\ell}) \setminus V(B_m)$ are sufficiently uniform relative to a large portion of the digraph coming from T_1 .

Lemma 6.7. *Assume the setup in Section 4 and suppose that $1/n \ll \varepsilon \ll 1/d$. With probability $1 - n^{-\omega(1)}$, all but $\varepsilon^5 n$ many $t \in T_2$ satisfy $\min(\deg_{B_{n-\ell}}^+(t, T_1), \deg_{B_{n-\ell}}^-(t, T_1)) \geq \sqrt{\log(1/\varepsilon)}$ and $\max(\deg_{B_{n-\ell}}^+(t), \deg_{B_{n-\ell}}^-(t)) \leq (\log(1/\varepsilon))^2$.*

Proof. The second part holds for all but at most $\varepsilon^5 n/3$ vertices immediately given the second item of Lemma 6.6 and basic computation.

Let us consider the number of total vertices in T_2 with at least 7 out-neighbors in T_2 , which is a set of size $\lfloor \varepsilon n \rfloor$. By Chernoff, it is easy to see that with probability $1 - n^{-\omega(1)}$, we have at most $\varepsilon^5 n / 3$ vertices with at least 7 out-neighbors in T_2 . The same holds for in-neighbors. We obtain a combined exceptional set in T_2 is of size at most $\varepsilon^5 n$. It suffices to check that for $t \in T_2$ not in this exceptional set, we have $\min(\deg_{B_{n-\ell}}^+(t, T_1), \deg_{B_{n-\ell}}^-(t, T_1)) \geq \sqrt{\log(1/\varepsilon)}$.

Now note that $B_{n-\ell}$ and A can be coupled so that $B_{n-\ell}$ is a sub-digraph of A *almost surely*. We can thus apply the first item of Lemma 6.6 (holding with probability $1 - n^{-\omega(1)}$).

Given this event, by the definition of T_2 as the highest ε^2 quantile (of minimum in- and out-degree within $B_{n-\ell}$) from a fixed set of size εn , we have with probability $1 - n^{-\omega(1)}$ that $\min(\deg_{B_{n-\ell}}^+(v), \deg_{B_{n-\ell}}^-(v)) \geq 2\sqrt{\log(1/\varepsilon)}$ for all $v \in T_2$. Since $\deg_{B_{n-\ell}}^+(v, T_1) \geq \deg_{B_{n-\ell}}^+(v) - 7 \geq \sqrt{\log(1/\varepsilon)}$ outside of our exceptional set, and the same holds for in-degrees, we are done. \square

Next we will require the following basic estimate regarding the maximum number of entries in a small diagonal block of the matrix.

Lemma 6.8. *With probability at least $1 - n^{-1+o(1)}$, for all $m \leq t \leq n$ and any S with $|S| \leq \sqrt{\log n}$ we have*

$$\sum_{i,j \in S} (B_t)_{ij} \leq |S|.$$

Proof. Noting the monotone nature of the estimate, it suffices to prove that

$$\sum_{i,j \in S} B_{ij} \leq |S|$$

for all $|S| \leq \sqrt{\log n}$. We have

$$\begin{aligned} \mathbb{P}\left(\exists S \subseteq [n]: |S| \leq \sqrt{\log n}, \sum_{i,j \in S} B_{ij} \geq |S| + 1\right) &\leq \sum_{k \leq \sqrt{\log n}} \binom{n}{k} \binom{k^2}{k+1} \left(\frac{\sqrt{\log n}}{n}\right)^{k+1} \\ &\leq \sum_{k \leq \sqrt{\log n}} \left(\frac{en}{k}\right)^k (ek)^{k+1} \left(\frac{\sqrt{\log n}}{n}\right)^{k+1} \\ &\leq \sum_{k \leq \sqrt{\log n}} k(e^2 \sqrt{\log n})^{k+1} n^{-1}, \end{aligned}$$

hence this failure probability is bounded by $n^{-1+o(1)}$ as desired. \square

We will also require the following variant event which handles sets that are substantially larger.

Lemma 6.9. *Let A be as in Section 4. There exists an absolute constant $C > 0$ such that the following holds. Let $1 \leq t \leq s \leq n \exp(-Cd)$. Then with probability at least $1 - e^{-s} n^{-1/2}$, we have*

$$\sup_{\substack{|S|=s \\ |T|=t \\ S \cap T = \emptyset}} \sum_{\substack{i \in S \\ j \in S \cup T}} A_{ij} \leq s + t + \frac{Cs \log(2d)}{\log(n/s)}.$$

Proof. Let $g(s) = \left\lfloor \frac{Cs \log(2d)}{\log(n/s)} \right\rfloor$. We have

$$\mathbb{P}\left[\sup_{\substack{|S|=s \\ |T|=t \\ S \cap T = \emptyset}} \sum_{\substack{i \in S \\ j \in S \cup T}} A_{ij} \geq s + t + g(s) + 1\right] \leq \binom{n}{s} \binom{n}{t} \mathbb{P}\left[\sum_{\substack{i \in [s] \\ j \in [s+t]}} A_{ij} \geq s + t + g(s) + 1\right]$$

$$\begin{aligned}
&\leq \left(\frac{en}{s}\right)^s \left(\frac{en}{t}\right)^t \binom{s(s+t)}{s+t+g(s)+1} \left(\frac{d}{n}\right)^{s+t+g(s)+1} \\
&\leq \left(\frac{en}{s}\right)^s \left(\frac{en}{t}\right)^t (es)^{s+t+g(s)+1} \left(\frac{d}{n}\right)^{s+t+g(s)+1} \\
&\leq (ed)^{4(s+t)} n^{-g(s)-1} s^{-s} t^{-t} s^s s^t s^{g(s)+1} \\
&\leq (ed)^{8s} (s/t)^t (s/n)^{g(s)+1} \leq e^s (ed)^{8s} (s/n)^{g(s)+1} \\
&\leq e^{-s} n^{-1/2}.
\end{aligned}$$

The final inequality is trivial to check when $g(s) \neq 0$ and when $g(s) = 0$, $(3d)^{40s} \leq n$ and hence the desired inequality also holds. \square

6.4. Unstructuredness events. We now define the set of events which will be used to guarantee that the kernel vectors to B_t and B_t^* are unstructured. We consider the hierarchy of constants

$$\varepsilon \ll \kappa \ll 1/d$$

where κ will control the size of the support of set; recall $\varepsilon = \mathbb{P}(\text{Pois}(d) \geq \Delta)$ and at the end of the argument we will take $\Delta \rightarrow \infty$ slowly. (At various points we will use that κ is small in terms of d .)

We say a $\{0, 1\}$ -matrix M is in \mathcal{D} if the degree sequence associated to M is $(d, \varepsilon^{1/2}, 16)$ -regular (Definition 6.3) and for all sets S of size at most $\sqrt{\log n}$ we have $\sum_{i,j \in S} M_{ij} \leq |S|$.

We say a $\{0, 1\}$ -matrix M is in $\mathcal{U}(r)$ if for all sets S of size in $[r, \kappa n]$ we have

$$|U(S)| \geq \alpha(|S|)|S|.$$

Finally, we say that a $\{0, 1\}$ -matrix M of with row indices T and column indices $T \cup \{t\}$ for $t \notin T$ is in \mathcal{U}^* if all column subsets $S \subseteq T \cup \{t\}$ containing t satisfy

$$U(S) \geq 1.$$

We now show that our random matrices B_t (and B_t^\dagger , and also B_t^*) are in unstructuredness sets $\mathcal{U}(r)$ and \mathcal{D} for various choices of r and t .

Lemma 6.10. *We have*

$$\mathbb{P}\left(\bigcap_{m \leq t \leq n} \{B_t^\dagger \in \mathcal{D}\} \cap \bigcap_{m \leq t \leq n} \{B_t \in \mathcal{D}\}\right) \geq 1 - n^{-1+o(1)}.$$

Proof. That the degree sequence of B is $(d, n^{-1/4}, 8)$ -regular follows from Lemma 6.6, noting that A and B can be coupled to differ in at most $(\log n)^3$ entries with probability $1 - n^{-\omega(1)}$, and noting that the maximum degree of B is bounded by $\log n$ with probability $1 - n^{-\omega(1)}$ (which allows control of the second bullet point of Definition 6.3 for small S).

Noting that B_t is obtained from B by removing at most $\varepsilon n + \ell$ vertices and the largest εn vertices have total in- and out-degree bounded by $O(\varepsilon(\log(1/\varepsilon) + d)n)$ from regularity of the degree sequence of B , the desired $(d, \varepsilon^{1/2}, 16)$ -regularity for all B_t follows immediately since $\varepsilon \ll 1/d$. This establishes the first part of \mathcal{D} for all relevant matrices.

The second part of the definition of \mathcal{D} holds for all our matrices with probability $1 - n^{-1+o(1)}$ due to Lemma 6.8. \square

Lemma 6.11. *We have*

$$\mathbb{P}\left(\bigcap_{m \leq t \leq n-\ell} \{B_t^\dagger \in \mathcal{U}((\log n)^{3/2})\} \cap \bigcap_{m \leq t \leq n-\ell} B_t \in \{\mathcal{U}((\log n)^{3/2})\}\right) \geq 1 - n^{-\omega(1)}.$$

Proof. It suffices to prove that for any $m \leq t \leq n - \ell$, we have

$$\mathbb{P}(B_t \in \mathcal{U}((\log n)^{3/2})) \geq 1 - n^{-\omega(1)};$$

the case of B_t^\dagger follows by symmetry. Note that B_t has a $(d, \varepsilon^{1/2}, 16)$ -regular degree sequence with probability $1 - n^{-\omega(1)}$. Therefore by Lemma 6.4 applied with $c = c(\min(d-1, 1/d), 16)$, it follows that the probability there exists a set S of k with $|\sum_{v \in S} \deg_{B_t}^+(v) - k| \leq k/\sqrt{\log(n/k)}$ and $|U(S)| \leq \alpha(k)k$ is at most $c^{-1} \exp(-ck) + n^{-\omega(1)}$ by Markov's inequality. We can take a union bound over $k \geq (\log n)^{3/2}$ to handle these S .

We now handle S such that $|\sum_{v \in S} \deg_{B_t}^+(v) - k| \geq k/\sqrt{\log(n/k)}$. If $\sum_{i,j \in S} (B_t)_{ij} \leq k - k/(4\sqrt{\log(n/k)})$, note that

$$|S \cap U(S)| \geq |S| - \sum_{i,j \in S} (B_t)_{ij} \geq k/(4\sqrt{\log(n/k)}) \geq k\alpha(k).$$

Thus we may restrict to sets S such that $\sum_{v \in S} \deg_{B_t}^+(v) \geq k + k/\sqrt{\log(n/k)}$ and $\sum_{i,j \in S} (B_t)_{ij} \geq k - k/(4\sqrt{\log(n/k)})$.

Further assuming that $U(S) \leq \alpha(k)k$ occurs, there exists a set T such that $|T| \leq 3k/(8\sqrt{\log(n/k)})$ with $T \cap S = \emptyset$ such that

$$\sum_{\substack{i \in S \\ j \in S \cup T}} (B_t)_{i,j} \geq k + \frac{k}{2\sqrt{\log(n/k)}}.$$

In particular, T can be taken to be the set of vertices outside S with at least 2 neighbors in S , truncating T to the appropriate size if it is too large. This contradicts Lemma 6.9 (which holds with probability $\geq 1 - e^{-k}n^{-1/2}$) since

$$k + |T| + \frac{Ck \log(2d)}{\log(n/k)} < k + \frac{k}{2\sqrt{\log(n/k)}}$$

with C as in Lemma 6.9 and since $\kappa \ll 1/d$. □

Lemma 6.12. *We have*

$$\mathbb{P}\left(\bigcap_{n-\ell \leq t \leq n} \{B_t^\dagger \in \mathcal{U}((\log \log n)^2)\} \cap \bigcap_{n-\ell \leq t \leq n} \{B_t^* \in \mathcal{U}((\log \log n)^2)\}\right) \geq 1 - (\log n)^{-\omega(1)}.$$

Proof. The result follows immediately from Lemma 6.1 and taking the union bound over sizes larger than $(\log \log n)^2$ if B_t has independent $\text{Ber}(d/n)$ entries. We can couple to the correct model with a polynomial loss in TV-distance by the remarks after Lemma 5.1. □

Lemma 6.13. *We have*

$$\mathbb{P}\left(\bigcap_{n-\ell \leq t \leq n} \{B_t^* \in \mathcal{U}^*\}\right) \geq 1 - (\log n)^{-\omega(1)}.$$

Proof. Unique-neighbor expansion of sets for size larger than $(\log \log n)^2$ follows immediately from Lemma 6.12. The remaining result follows immediately from Lemma 6.2 and the remarks following Lemma 5.1 to account for slight differences in the random model. □

7. SPREADNESS OF NEAR KERNEL VECTORS

We now extract the crucial vector spreadness estimates for our results, Propositions 7.1, 7.2, and 7.3. Recall the digraph unstructuredness events $\mathcal{U}(r), \mathcal{D}, \mathcal{U}^*$. Proposition 7.1 will handle the first epoch.

Proposition 7.1. *Suppose that $\varepsilon \ll 1/d$ and fix $z \in \mathbb{C} \setminus \{0\}$. There are constants $c = c(d) > 0$ and $C' = C'(d, z) > 0$ so that the following holds. Let $m \leq t \leq n - \ell$ and $M \in \{B_t, B_t^\dagger\}$. Suppose that $M \in \mathcal{U}((\log n)^{3/2})$ and $M \in \mathcal{D}$. Whenever $(\log n)^{7/4} \leq k \leq n$ and $v \in \mathbb{C}^t$ is a vector such that $v_k^* \geq (k/t)^2/\sqrt{t}$ and*

$$\|(M - zI_t)v\|_2 \leq \exp(-C'(\log(2n/k))^6),$$

we must have

$$\sup_{\theta \in \mathbb{R}} \#\left\{i: |v_i - \theta| \leq \frac{\exp(-C'(\log(2n/k))^7)}{\sqrt{n}}\right\} \leq (1-c)n.$$

Propositions 7.2 and 7.3 will handle the second epoch. The proofs of these are similar except the proof of Proposition 7.3 is strictly more complicated, so we omit the proof of Proposition 7.2.

Proposition 7.2. *Suppose $\varepsilon \ll 1/d$ and fix $z \in \mathbb{C} \setminus \{0\}$. There are constants $C' = C'(d, z) > 0$ and $c = c(d) > 0$ such that the following holds. Consider $n - \ell \leq t \leq n - 1$ and let $M = B_t^\dagger$. Suppose $M \in \mathcal{U}((\log \log n)^2)$ and $M \in \mathcal{D}$. Then whenever $v \in \mathbb{C}^t$ is a unit vector and*

$$\|(M - zI_t)v\|_2 \leq \exp(-C'(\log n)^6),$$

we must have

$$v_{\lfloor cn \rfloor}^* \geq \frac{\exp(-C'(\log n)^7)}{\sqrt{n}}.$$

Proposition 7.3. *Suppose $\varepsilon \ll 1/d$ and fix $z \in \mathbb{C} \setminus \{0\}$ such that $|z| \neq 1$. There are constants $C' = C'(d, z) > 0$ and $c = c(d) > 0$ such that the following holds. Consider $n - \ell \leq t \leq n - 1$ and let $M = B_t^*$. Suppose $M \in \mathcal{U}((\log \log n)^2)$ and $M \in \mathcal{D}$. Additionally, suppose that $M \in \mathcal{U}^*$. Then whenever $v \in \mathbb{C}^t$ is a unit vector and*

$$\|(M - zI_{(t-1) \times t})v\|_2 \leq \exp(-C'(\log n)^6),$$

we must have

$$v_{\lfloor cn \rfloor}^* \geq \frac{\exp(-C'(\log n)^7)}{\sqrt{n}}.$$

7.1. Initial estimates and setup. We first require the connection between unique neighborhood expansion and the images of vectors.

Observation 7.4. *Let M be a $(t-1) \times t$ or $t \times t$ dimensional $\{0, 1\}$ -matrix. For $\ell \leq t$, let $v \in \mathbb{C}^t$ and let S be the set of the ℓ largest coordinates of v in absolute value. Then $|((M - zI)v_S)_i| \geq v_\ell^* \min(|z|, 1)$ for all $i \in U(S)$ where $U(S)$ is defined with respect to the matrix M and I is the identity matrix with dimensions corresponding to M .*

Proof. We consider two cases. If $i \in U(S) \setminus S$ there is unique $j \in S$ with $(M - zI)_{ij} \neq 0$. As M is a $\{0, 1\}$ -matrix, we additionally have $M_{ij} = 1$ and we have $|((M - zI)v_S)_i| = |v_j| \geq v_\ell^*$. This proves the observation in this case.

For $i \in U(S) \cap S$ we have $M_{ij} = 0$ for all $j \in S$. So

$$|((M - zI)v_S)_i| = |(M - zI)_{ii}v_i| = |(-z)v_i| = |z||v_i| \geq v_\ell^*|z|,$$

which proves the observation. \square

Recall that above we defined the function $\alpha(x) = (\log(n/x))^{-2}$. Here we define the function

$$(15) \quad g(x) = \left\lceil \frac{\alpha(x)x}{2^{15}(d + \log(n/x))} \right\rceil.$$

We require the following trivial iteration lemma.

Lemma 7.5. *Let $k_0 = k$ and define*

$$k_i = k_{i-1} + g(k_{i-1}),$$

for all $i \geq 1$. Let τ be the minimal value such that $k_\tau \geq n/2$. Then $\tau \leq 2^{17}d(\log(n/k))^4$.

Proof. For $k \geq n/2$, the result is trivial. Furthermore note that it takes at most $2^{15}(d + \log(n/k))/\alpha(k) \leq 2^{16}d(\log(n/k))^3$ steps to double. As there are at most $2 \log(n/k)$ doublings required, the desired result follows immediately. \square

We finally will require the following graph-theoretic estimate which will allow us to eliminate graphs with extremely large level sets.

Lemma 7.6. *Consider an $\ell \times m$ dimensional $\{0, 1\}$ -matrix M with $\ell \in \{m-1, m\}$, $z \neq 0$, and let $\theta > 0$. Suppose that M has at least $e^{-d}n/2$ vertices with in-degree zero (i.e., zero rows). Then for a unit vector $v \in \mathbb{C}^m$ such that*

$$\|(M - zI_{\ell \times m})v\| \leq \theta|z|,$$

there are at least $e^{-d}n/4$ indices j such that

$$|v_j| \leq 2e^{d/2}\theta n^{-1/2}.$$

Proof. Note that for any index $i \in [\ell]$ such that $\deg_M^-(i) = 0$, we have

$$((M - zI_{\ell \times m})v)_i = -zv_i.$$

Therefore since $\|(M - zI_{\ell \times m})v\| \leq \theta|z|$, we have

$$\sum_{\deg_M^-(i)=0} |zv_i|^2 \leq \theta^2|z|^2.$$

Applying Markov's inequality we derive the desired conclusion. \square

7.2. Unstructured almost-kernel vectors for the first epoch. We are now in position to prove Proposition 7.1.

Proof of Proposition 7.1. As $M \in \mathcal{D}$, we have that M has at least $e^{-d}n/2$ vertices with out-degree zero. Therefore by Lemma 7.6, we have that

$$|\{v_i \mid |v_i| \leq 2e^{d/2}|z|^{-1} \exp(-C'(\log(2n/k))^6)/\sqrt{n}\}| \geq e^{-d}n/4.$$

Taking C' sufficiently large, we have that for all $|\theta| \geq 2 \exp(-C'(\log(2n/k))^7)/\sqrt{n}$ that

$$\#\left\{i: |v_i - \theta| \leq \frac{\exp(-C'(\log(2n/k))^7)}{\sqrt{n}}\right\} \geq e^{-d}n/4.$$

Thus it suffices to prove that

$$\#\left\{i: |v_i| \leq \frac{4 \exp(-C'(\log(2n/k))^6)}{\sqrt{n}}\right\} \leq (1-c)n.$$

By assumption, we have that $v_k^* \geq (k/t)^2/\sqrt{t}$. We claim that it suffices to prove for all $k' \in [k, cn]$ that

$$v_{k'+g(k')}^* \geq v_{k'}^* \cdot \left(\frac{k'}{dn \min(|z|, |z|^{-1})}\right)^2.$$

This immediately implies the desired result since then

$$v_{cn}^* \geq v_k^* \cdot \prod_{i=1}^{\tau} \left(\frac{k_i}{dn \min(|z|, |z|^{-1})}\right)^2 \geq \exp(-C' \log(n/k)^5)/\sqrt{n},$$

choosing C' sufficiently large and defining k_i, τ with bounds as in Lemma 7.5.

To prove the claim, suppose that

$$v_{k'+g(k')}^* < v_{k'}^* \cdot \left(\frac{k'}{dn \min(|z|, |z|^{-1})} \right)^2$$

and let k' is chosen to be minimal such value in $[k, cn]$. Let S denote the indices of the largest k' coordinates of v and S' denote the indices of the coordinates with magnitude from the $(k'+1)$ st largest to the $(k'+g(k'))$ th largest. By Observation 7.4, for all $j \in U(S)$ we have

$$|((M - zI)v_S)_j| \geq v_{k'}^* \min(|z|, 1).$$

Since $M \in \mathcal{U}((\log n)^{3/2})$, we have $|U(S)| \geq \alpha(k')k'$. Since $M \in \mathcal{D}$, by the second item of Definition 6.3, we have that there are at most

$$16(d + \log(m/g(k')))g(k') \leq \alpha(k')k'/4$$

neighbors of S' . Furthermore by the second item of Definition 6.3, for $t \geq 32d$ there are at most $\exp(-t/32)n$ vertices of in-degree larger than t . In particular, there are at most $\alpha(k')k'/4$ of in-degree larger than n/k' assuming that c is a sufficiently small function of d .

Therefore, define a row index j in $U(S)$ to be suitable if it has in-degree bounded by n/k' and is not adjacent to any column index in S' . Note we have proven that there are at least $\alpha(k')k'/2$ suitable indices. For each suitable index j , we have that

$$\begin{aligned} |((M - zI)v)_j| &= |((M - zI)(v_S + v_{S'} + v_{([t] \setminus (S \cup S'))})_j)| \\ &= |((M - zI)(v_S + v_{([t] \setminus (S \cup S'))})_j)| \\ &\geq v_{k'}^* \min(|z|, 1) - \left(\frac{n}{k'} + |z| \right) v_{k'+g(k')}^* \geq v_{k'}^* \min(|z|, 1)/2. \end{aligned}$$

This gives us a contradiction as $v_{k'}^* \geq \exp(-C'(\log(2n/k))^5)/\sqrt{n}$ and therefore

$$\begin{aligned} \|((M - zI)v)\|_2 &\geq (\alpha(k')k'/2)^{1/2} \cdot \exp(-C'(\log(2n/k))^5) \min(|z|, 1)/(2\sqrt{n}) \\ &\geq \exp(-C' \log(n/k)^6) \end{aligned}$$

since C' is sufficiently large as a function of d and z . \square

7.3. Unstructured almost-kernel vectors for the second epoch. In order to deal with the cases $t \geq n - \ell$, we split into two cases. For large support almost-kernel vectors, we use a similar argument to the proof of Proposition 7.1. On the other hand, small support almost-kernel vectors must essentially be kernel vectors of circulant matrices, which we can explicitly handle using the following Lemma 7.7.

Lemma 7.7. *Let Y be the $s \times s$ matrix where $Y_{ij} = 1$ if and only if $i \equiv j + 1 \pmod{s}$. Then*

$$\sigma_s(Y - zI_s) \geq |z^s - 1|/(|z| + 1)^{s-1}.$$

Proof. Notice that by direct computation that

$$\det((Y - zI_s)^\dagger(Y - zI_s)) = |z^s - 1|^2$$

and that

$$\sigma_1(Y - zI_s) \leq |z| + 1.$$

The desired result follows from

$$\det((Y - zI_s)^\dagger(Y - zI_s)) \leq \sigma_s(Y - zI_s)^2 \cdot \sigma_1(Y - zI_s)^{2(s-1)}$$

and dividing. \square

Proof of Proposition 7.3. It suffices to prove that $v_{k'+1}^* \geq v_{k'}^*/(C'n)$ for $k' \leq (\log \log n)^3$. This is sufficient since after this point a modification of the argument in Proposition 7.1 easily completes the proof.

Otherwise suppose that $v_{k'+1}^* \leq v_{k'}^*/(C'n)$ for a minimal such $k' \leq (\log \log n)^3$, and note that since v is a unit vector we have $v_1^* \geq 1/\sqrt{n}$. Let S be the set of indices corresponding to the k' largest coordinates. Furthermore notice that the in- and out-degree of every vertex is bounded by $(\log n)^2$ since $M \in \mathcal{D}$. If $|U(S)| \geq 1$, for any $j \in U(S)$ we have

$$|((M - zI_{(t-1) \times t})v)_j| \geq v_{k'}^* \cdot \min(|z|, 1) - ((\log n)^2 + |z|)v_{k'+1}^* \geq v_{k'}^* \cdot \min(|z|, 1)/2.$$

Notice that $v_{k'}^* \geq \exp(-C'(\log n)^2)$ which provides the desired contradiction in this case.

Therefore we may assume that $U(S) = \emptyset$. By \mathcal{U}^* , we deduce that $t \notin S$. Furthermore, since $t \notin S$ and $U(S) = \emptyset$, every vertex in S has at least 1 in-neighbor from S . By the final condition of $M \in \mathbb{D}$, we have

$$\sum_{i,j \in S} M_{ij} \leq |S|$$

and thus every vertex in S has exactly 1 in-neighbor from S . We refine S as follows. If there is no vertex in the current set with out-degree zero terminate; else remove a vertex of out-degree zero and iterate. This process terminates with a set T in which the induced directed subgraph is exactly a collection of cycles (possibly of length 1). Furthermore for any vertex in T , we have that it has no in-neighbor from $S \setminus T$. Adding $-z$ to the diagonal entries of the adjacency matrix of the induced digraph $M[T]$, we get a disjoint collection of circulant matrices of exactly the form in Lemma 7.7. This argument uses crucially that $t \notin S$ and therefore $t \notin T$.

Finally, applying Lemma 7.7 we have, writing $s = (\log \log n)^3$,

$$\begin{aligned} \|(M - zI_{(t-1) \times t})v\|_2 &\geq \|((M - zI_{(t-1) \times t})v)_T\|_2 \\ &= \|((M - zI_{(t-1) \times t})v_S)_T\|_2 - \sqrt{n}((\log n)^2 + |z|)v_{k'+1}^* \\ &= \|((M - zI_{(t-1) \times t})v_T)_T\|_2 - \sqrt{n}((\log n)^2 + |z|)v_{k'+1}^* \\ &\geq \|v_T\|_2 \cdot \min\{|z^t - 1|/(|z| + 1)^{t-1} : t \in [s]\} - \sqrt{n}((\log n)^2 + |z|)v_{k'+1}^* \\ &\geq v_{k'}^* \cdot \|z\| - 1/(|z| + 1)^{s-1} - \sqrt{n}((\log n)^2 + |z|)v_{k'+1}^* \\ &\geq v_{k'}^* \cdot \|z\| - 1 \exp(-C'(\log \log n)^4) - \sqrt{n}((\log n)^2 + |z|)v_{k'+1}^* \\ &\geq \exp(-C'(\log n)^2) \end{aligned}$$

given that C' is a sufficiently large function of z and d . This contradicts the assumption in the lemma statement and completes the proof. \square

8. ANTICONCENTRATION ESTIMATES

We will require the following cutoff parameters; note that in the definition below the constant $K = K(d, z)$ will be chosen later. We define

$$\varepsilon_r = \exp(-K(\log(n/r))^9).$$

We will now consider the result of adding in vertex v_t in our walk. From now on we will use the common abuse of writing e_t for the column identity vector which is supported only on the index corresponding to v_t . (In fact, we may relabel so that v_t is labeled t if we so desire, since conjugation by a permutation matrix does not change the spectrum nor the singular values.)

We first prove that the projection onto the bottom set of singular vectors is unlikely to be extremely small; we initially restrict to the first epoch.

Lemma 8.1. *There exists $C = C(d) > 0$ such that the following holds. Fix $z \neq 0$ and let $m + 1 \leq t \leq n - \ell$ and $r \geq (\log n)^{5/3}$. Let $B_{t-1}^\dagger, B_{t-1} \in \mathcal{U}((\log n)^{3/2}) \cap \mathcal{D}$ and suppose that*

$$\min(\deg_{B_{n-\ell}}^+(v_t, T_1), \deg_{B_{n-\ell}}^-(v_t, T_1)) \geq \sqrt{\log(1/\varepsilon)}$$

(which is measurable given \mathcal{F}_t').

Define $P_{r,M}$ to denote the projection onto the bottom r right-singular vectors of M . If $\sigma_{(t-1)-r/2}(B_{t-1} - zI_{t-1,t-1}) \leq \varepsilon_r$, then for all $u \in \mathbb{C}^{t-1}$,

$$\mathbb{P}(\|P_{r, B_{t-1}^\dagger - zI_{t-1}}(B_t^* e_t + u)\|_2 < \varepsilon_r | \mathcal{F}_t') \leq C(\log(1/\varepsilon))^{-1/4}.$$

Furthermore for all $u \in \mathbb{C}^t$,

$$\mathbb{P}(\|P_{r, B_t^* - zI_{(t-1) \times t}}(B_t^\dagger e_t + u)\|_2 < \varepsilon_r | \mathcal{F}_t', B_t^*) \leq C(\log(1/\varepsilon))^{-1/4}.$$

The second lemma handles the analogous projection onto the bottom singular vector (which is all that is needed) in the second epoch.

Lemma 8.2. *There exists $C = C(d) > 0$ such that the following holds. Let $|z| \neq 0, 1$ and $n - \ell + 1 \leq t \leq n$. Define $P_{r,M}$ to denote the projection onto the bottom r right singular vectors of M .*

If $\sigma_{t-1}(B_{t-1} - zI_{t-1}) \leq \varepsilon_1$, $B_{t-1}^\dagger \in \mathcal{U}((\log \log n)^2)$ and $B_{t-1}^\dagger \in \mathcal{D}$, then for all $u \in \mathbb{C}^t$,

$$\mathbb{P}(\|P_{1, B_{t-1}^\dagger - zI_{t-1}}(B_t^* e_t + u)\|_2 < \varepsilon_1 | \mathcal{F}_t) \leq C(\log n)^{-1/4}.$$

Furthermore if $B_t^* \in \mathcal{U}((\log \log n)^2)$, $B_t^* \in \mathcal{U}^*$ and $B_{t-1}^\dagger \in \mathcal{D}$ then for all $u \in \mathbb{C}^t$,

$$\mathbb{P}(\|P_{1, B_t^* - zI_{(t-1) \times t}}(B_t^\dagger e_t + u)\|_2 < \varepsilon_1 | \mathcal{F}_t, B_t^*) \leq C(\log n)^{-1/4}.$$

These projection inequalities are designed to complement the crucial linear algebra input in this paper which is Proposition 10.3.

8.1. Anticoncentration estimates against a fixed vector. We will first require the Lévy concentration function. For a (real or complex) random variable Γ ,

$$\mathcal{L}(\Gamma, t) = \sup_{z \in \mathbb{C}} \mathbb{P}(|\Gamma - z| \leq t).$$

We will require the following anticoncentration inequality due to Kolmogorov–Lévy–Rogozin [18, 24] (see e.g. [25, Lemma 3.2]).

Lemma 8.3. *There exists an absolute constant $C > 0$ such that the following holds. Let ξ_1, \dots, ξ_n be independent real or complex random variables. Then, for any real numbers $r > 0$, we have*

$$\mathcal{L}\left(\sum_{i=1}^n \xi_i, r\right) \leq \frac{C}{\sqrt{\sum_{i=1}^n (1 - \mathcal{L}(\xi_i, r))}}.$$

We will also require the following ‘‘slice’’ anticoncentration inequality; the proof is essentially a quantitative version of the [11, Lemma 4.2].

Lemma 8.4. *There exists $C > 0$ such that the following holds. Fix $\delta, \gamma > 0$ and sample $\xi \in \{0, 1\}^n$ uniformly at random such that $\sum_{i=1}^n \xi_i = m$ and $1 \leq m \leq n/2$. Let $v \in \mathbb{C}^n$ such that*

$$\sup_{\theta \in \mathbb{C}} \#\{i: |v_i - \theta| \leq \delta\} \leq (1 - \gamma)n.$$

Then

$$\mathcal{L}\left(\sum_{i=1}^n \xi_i v_i, \delta\right) \leq C\left((\gamma m)^{-1/2} + \exp(-C^{-1}\gamma^2 m)\right).$$

Proof. Let π be a uniformly random injective function $\{1, \dots, 2m\} \rightarrow \{1, \dots, n\}$ and let $\mathbf{x} = (x_1, \dots, x_m)$ be a sequence of independent $\text{Ber}(1/2)$ random variables. Then choose the positions of the m ones in ξ as follows. For each $i \in \{1, \dots, m\}$, if $x_i = 0$ we set $\xi_{\pi(i)} = 1$ and $\xi_{\pi(i+m)} = 0$, and if $x_i = 1$ let $\xi_{\pi(i+m)} = 1$ and $\xi_{\pi(i)} = 0$. All other indices of ξ are set to 0. It is clear that ξ has the correct distribution.

We call an index i separated if

$$|v_{\pi(i)} - v_{\pi(i+m)}| \geq \delta$$

and let Y denote the number of separated indices, which is dependent only on π . Note that

$$\#\{(i, j): |v_i - v_j| \geq \delta\} \geq \gamma n^2/2,$$

else there exists an index i such that

$$\#\{j: |v_i - v_j| \leq \delta\} \geq (1 - \gamma)n$$

and taking $\theta = v_i$ we have violated our assumption. Thus $\mathbb{E}Y \gtrsim \gamma m$ and by applying Lemma A.3 we easily see $\mathbb{P}(Y \leq \mathbb{E}Y/2) \leq \exp(-\Omega(\gamma^2 m))$.

Therefore,

$$\begin{aligned} & \sup_{\theta \in \mathbb{C}} \mathbb{P}\left(\left|\sum_{i=1}^n \xi_i v_i - \theta\right| \leq \delta\right) \\ & \leq \mathbb{P}(Y \leq \mathbb{E}Y/2) + \sup_{\theta, \pi} \mathbb{1}_{Y \geq \mathbb{E}Y/2} \mathbb{P}\left(\left|\sum_{i=1}^m \xi_{\pi(i)} v_{\pi(i)} + \xi_{\pi(i+m)} v_{\pi(i+m)} - \theta\right| \leq \delta \mid \pi\right) \\ & \leq \mathbb{P}(Y \leq \mathbb{E}Y/2) + \sup_{\theta, \pi} \mathbb{1}_{Y \geq \mathbb{E}Y/2} \mathbb{P}\left(\left|\sum_{i=1}^m (v_{\pi(i)} + x_i(v_{\pi(i+m)} - v_{\pi(i)})) - \theta\right| \leq \delta \mid \pi\right) \\ & \lesssim \exp(-\Omega(\gamma^2 m)) + (\gamma m)^{-1/2} \end{aligned}$$

where we have used that $(x_i)_{i \in [m]}$ is distributed as $\text{Ber}(1/2)^{\otimes m}$ given π and then applied Lemma 8.3 to the set of separated i where $|v_{\pi(i+m)} - v_{\pi(i)}| \geq \delta$ (i.e., those counted by $Y \geq \mathbb{E}Y/2 \gtrsim \gamma m$). \square

8.2. Existence of a well-balanced basis. We require the existence of a well-balanced basis for the span of the least singular vectors. The following lemma of Litvak, Lytova, Tikhomirov, Tomczak-Jaegermann, and Youssef [19, Lemma 4.3] gives a decent basis for any vector space.

Lemma 8.5. *Let $V \subseteq \mathbb{C}^n$ be a k -dimensional \mathbb{C} -vector space. There exists an orthonormal basis B of V so that for all $v \in B$, we have $v_{c_k}^* \geq ck^{1/2}n^{-1}$, where $c > 0$ is an absolute constant.*

8.3. Proof of Lemma 8.1. We are now in position to prove Lemma 8.1.

Proof of Lemma 8.1. We only prove the second item of the lemma; the first item is strictly simpler. Recall that B_t^* is a $(t-1) \times t$ matrix and note that $B_t^\dagger e_t$ is the Hermitian conjugate of the row added to B_t^* to obtain B_t^\dagger . By assumption we have $\sigma_{(t-1)-r/2}(B_{t-1} - zI_{t-1}) \leq \varepsilon_r$. By Cauchy interlacing applied for singular values (see e.g. Fact 10.1) we have that $\sigma_{t-r/2}(B_t^* - zI_{(t-1) \times t}) \leq \varepsilon_r$.

Therefore there exists a vector space $W \subseteq \mathbb{C}^t$ of dimension $r/2$ such that for unit vectors v in W we have $\|(B_t^* - zI_{(t-1) \times t})v\|_2 \leq \varepsilon_r$. Let $W' = W \cap \{v_t = 0: v \in \mathbb{C}^t\}$ and note that the dimension of W' is at least $r/2 - 1 \geq r/4$. Letting $\pi: (x_1, \dots, x_t) \rightarrow (x_1, \dots, x_{t-1})$, we have for all unit vectors v in W' that

$$\|(B_t^* - zI_{(t-1) \times t})v\|_2 = \|(B_{t-1} - zI_{t-1})\pi(v)\|_2 \leq \varepsilon_r.$$

Applying Lemma 8.5, there exist unit vectors $w_1, \dots, w_{r/4}$ in W' such that

$$(w_j)_{cr}^* \geq cr^{1/2}n^{-1}$$

for all $j \leq r/4$ and an absolute constant $c > 0$. Since $\|(B_t^* - zI_{(t-1) \times t})w_j\|_2 = \|(B_{t-1} - zI_{t-1})w_j\|_2 \leq \varepsilon_r$, we may use Proposition 7.1. We have

$$\sup_{\theta \in \mathbb{C}} \#\{i: |(w_j)_i - \theta| \leq \exp(-C'(\log(n/r))^7)n^{-1/2}\} \leq (1 - c')n$$

for all $1 \leq j \leq r/4$, where c' is a constant depending only on d while C' is a function of d and z .

Note that $B_t^\dagger e_t$ given $\mathcal{F}'_t \cup B_t^*$ is deterministic on the indices outside of T_1 and on the indices T_1 is a uniformly random $\{0, 1\}$ -vectors with a fixed sum by Fact 4.5. Furthermore note that since $|T_1| = \lfloor n(1 - \varepsilon) \rfloor \geq n(1 - c'/4)$ (as $\varepsilon \ll 1/d$), we still have that the largest approximate level set of each w_j when restricted to T_1 occupies at most a $(1 - c'/2)$ -fraction.

Note that the fixed sum of $B_t^\dagger e_t$ on T_1 is at least $\sqrt{\log(1/\varepsilon)}$ by assumption and at most $(\log n)^2$ by the maximum degree assumption implicit in \mathcal{D} . Therefore for any deterministic vector u we have

$$\begin{aligned} \mathbb{P}(|\langle w_k, B_t^\dagger e_t + u \rangle| \leq \exp(-C'(\log(n/r))^7)n^{-1/2} | \mathcal{F}'_t, B_t) \\ \leq \sup_{\theta \in \mathbb{C}} \mathbb{P}(|\langle (w_k)_{T_1}, (B_t^\dagger e_t)_{T_1} \rangle - \theta| \leq \exp(-C'(\log(n/r))^7)n^{-1/2} | \mathcal{F}'_t, B_t) \lesssim (\log(1/\varepsilon))^{-1/4}, \end{aligned}$$

where the inequality follows from Lemma 8.4 (note that the implicit constant here depends only on c' in the size of the largest level set and hence only on d).

Note that for a set of nonnegative random variables X_1, \dots, X_k , by Markov's inequality

$$\mathbb{P}\left(\sum_{i=1}^k X_i \leq \tau\right) \leq \mathbb{P}(|\{i: X_i \leq 2\tau/k\}| \geq k/2) \leq \frac{2}{k} \mathbb{E}\left[\sum_{i=1}^k \mathbb{1}_{X_i \leq 2\tau/k}\right] \leq 2 \sup_{i \in [k]} \mathbb{P}(X_i \leq 2\tau/k).$$

Therefore

$$\begin{aligned} \mathbb{P}(\|P_{r, B_t^* - zI_{(t-1) \times t}}(B_t^\dagger e_t + u)\|_2 < \varepsilon_r | \mathcal{F}'_t \cup B_t^*) &\leq \mathbb{P}\left(\sum_{k=1}^{r/4} |\langle w_k, B_t^\dagger e_t + u \rangle|^2 < \varepsilon_r^2 | \mathcal{F}'_t, B_t^*\right) \\ &\leq \sup_k 2\mathbb{P}(|\langle w_k, B_t^\dagger e_t + u \rangle| < \varepsilon_r \cdot (r/8)^{-1/2} | \mathcal{F}'_t, B_t^*) \\ &\lesssim (\log(1/\varepsilon))^{-1/4}, \end{aligned}$$

which is precisely the desired result. Here we have used that the constant K defining ε_r is chosen as a sufficiently large function of d and z . \square

8.4. Proof of Lemma 8.2. We now prove Lemma 8.2; since the randomness in the second epoch is purely independent, the analysis simplifies substantially in this case.

Proof of Lemma 8.2. We will only prove the second case; the first is essentially identical except we use Proposition 7.2. Let w denote the least singular vector of $B_t^* - zI_{(t-1) \times t}$; as this is a $(t-1) \times t$ matrix we have $(B_t^* - zI_{(t-1) \times t})w = 0$. By Proposition 7.3, we have

$$w_{cn}^* \geq \exp(-C'(\log n)^7)n^{-1/2} =: \gamma$$

with $C' = C'(d, z) > 0$ and $c = c(d) > 0$. Therefore

$$\begin{aligned} \mathbb{P}(\|P_{1, B_t^* - zI_{(t-1) \times t}}(B_t^\dagger e_t + u)\|_2 < \varepsilon_1 | \mathcal{F}_t, B_t^*) &= \mathbb{P}(|\langle w, B_t^\dagger e_t + u \rangle| \leq \varepsilon_1 | \mathcal{F}_t, B_t^*) \\ &\leq \sup_{\theta \in \mathbb{C}} \mathbb{P}(|\langle w, B_t^\dagger e_t \rangle - \theta| \leq \varepsilon_1 | \mathcal{F}_t, B_t^*) \end{aligned}$$

$$\begin{aligned}
&= \sup_{\theta \in \mathbb{C}} \mathbb{P}_{\delta_j \sim \text{Ber}(\sqrt{\log n}/n)} \left(\left| \sum_{j=1}^t w_j \delta_j - \theta \right| \leq \varepsilon_1 \right) \\
&\leq \frac{1}{\sqrt{\sum_{j=1}^t (\sqrt{\log n}/n) \cdot \mathbb{1}_{|w_j| \geq \gamma}}} \lesssim (\log n)^{-1/4}
\end{aligned}$$

We have used Fact 4.1 to rewrite $\langle w, B_t^\dagger e_t \rangle$ as an sum of weighted independent random Bernoulli random variables and applied Lemma 8.3 with $r = \gamma$ and $\xi_i = w_i \delta_i$. The implicit constant in the final inequality comes from the implicit constant on the number of coordinates in w with size larger than γ , which depends only on d by Proposition 7.3. \square

9. RANDOM WALK LEMMA

We now prove the following simple probabilistic lemma which shows that if $(X_s)_{s=1}^k$ has sufficient downward drift, then X_k is small with sufficiently high probability. Results of this form originate in work of Costello, Tao, and Vu [7] on singularity of symmetric random matrices, and have been used to study singularity and rank in sparse random matrices [10, 13].

Lemma 9.1. *There exists an absolute constant $C > 0$ such that the following holds. Let $(\mathcal{F}_s)_{s=1}^k$ be a filtration and let $(X_s)_{s=1}^k$ be a sequence of random variables for which X_s is \mathcal{F}_s -measurable. Suppose that:*

- $X_1 \leq k/4$ almost surely;
- $X_{i+1} \leq X_i + 1$ almost surely;
- If $X_t \geq \lfloor (k-t)/8 \rfloor$, we have $\mathbb{P}(X_{t+1} \leq X_t - 1 + \mathbb{1}_{X_t=0} | \mathcal{F}_s) \geq 1 - p$.

Then for $t \geq 1$, we have

$$\mathbb{P}(X_k \geq t) \leq (Cp)^{t/2}.$$

Proof. We will choose C at the end of the proof sufficient large. We may assume that $p \leq C^{-1}$ as otherwise the result is vacuous. Furthermore we may assume that $k \geq 5$; for $k \leq 4$ we have $\mathbb{P}(X_k \geq t) \leq \binom{k}{t} p^t \leq 16p^{t/2}$.

Define $Y_t = (1/\sqrt{p})^{X_t} - 1$. We claim that

$$(16) \quad \mathbb{E}[Y_{t+1} | \mathcal{F}_t] \leq p^{-\lfloor (k-t)/8 \rfloor / 2} + (2\sqrt{p})Y_t + 2\sqrt{p} \leq 3p^{-\lfloor (k-t)/8 \rfloor / 2} + (2\sqrt{p})Y_t.$$

Indeed, notice that if $X_t < \lfloor (k-t)/8 \rfloor$ then we have $X_{t+1} \leq X_t + 1 \leq \lfloor (k-t)/8 \rfloor$ and the result follows immediately. Else if $X_t > 0$ and $X_t \geq \lfloor (k-t)/8 \rfloor$ then we have

$$\mathbb{E}[Y_{t+1} + 1 | \mathcal{F}_t] \leq \sqrt{p}(Y_t + 1) + p \cdot (1/\sqrt{p})(Y_t + 1),$$

which also implies (16). Finally if $X_t = 0$ and $X_t \geq \lfloor (k-t)/8 \rfloor$ then we have

$$\mathbb{E}[Y_{t+1} + 1 | \mathcal{F}_t] \leq (1/\sqrt{p} - 1)p \leq \sqrt{p},$$

which gives (16); thus we have verified (16) in all cases. Let $Z_t = p^{(k-t)/8} Y_t$ and note for $t \leq k-1$ that

$$\mathbb{E}[Z_{t+1} | \mathcal{F}_t] \leq p^{(k-t-1)/8} (3p^{-\lfloor (k-t)/8 \rfloor / 2} + (2\sqrt{p})Y_t) \leq 3 + p^{1/3} Z_t$$

if $p \leq C^{-1} \leq 1/2^{24}$. Iterating this bound we have

$$\begin{aligned}
\mathbb{E}Z_k &\leq 3 + 3 \cdot p^{1/3} + 3 \cdot (p^{1/3})^2 + \dots + p^{(k-1)/3} \mathbb{E}Z_1 \\
&\leq 4 + p^{(k-1)/3} p^{(k-1)/8} p^{-k/8} \leq 5
\end{aligned}$$

as $p \leq C^{-1}$ and $k \geq 5$. By Markov's inequality, for $p \leq C^{-1}$ we have

$$\mathbb{P}(X_k \geq t) = \mathbb{P}(Z_k \geq p^{-t/2} - 1) \leq 6p^{t/2},$$

as desired. \square

10. SINGULAR VALUE UPDATE FORMULA

The goal of this section is to prove the key singular value update formula Proposition 10.3 which will allow us to gain control of the least singular values of our final matrix by passing control along the random walk. This section is self-contained and we do not adopt any of the global variables that are present in the other parts of the paper. We first require Cauchy interlacing for singular values.

Fact 10.1. *Let M be an $n \times m$ matrix and let M' be M with a row added. Then*

$$\sigma_m(M) \leq \sigma_m(M') \leq \sigma_{m-1}(M) \leq \sigma_{m-1}(M') \leq \cdots \leq \sigma_1(M) \leq \sigma_1(M').$$

Next we require the following basic fact from linear algebra.

Fact 10.2. *Let M be an $n \times m$ matrix with $n \leq m$. We have $\sigma_i(M) = \sigma_i(M^\dagger)$ for $i \leq n$ and $\sigma_i(M) = 0$ for $n + 1 \leq i \leq m$.*

Finally we require a lemma which relates the product of a given segment of singular values of M to M' . Note that unlike the corresponding lemma [27, Lemma 7.2] in our companion paper, the number of singular values considered is the same on both sides. We make progress as we move our chunk of singular values “closer” in index to the minimal singular values for M' .

Proposition 10.3. *Let $n \leq m$, let M be an $n \times m$ matrix, and let M' be an $(n + 1) \times m$ matrix obtained by adding the row X to M . For $1 \leq k - 1 \leq \ell < m$, we have*

$$\prod_{i=k}^{\ell+1} \sigma_i(M') \geq \|PX^\dagger\|_2 \cdot (\|X\|_2^2 + \sigma_{k-1}(M)^2)^{-1/2} \cdot \prod_{i=k-1}^{\ell} \sigma_i(M),$$

where P is the orthogonal projection onto the span of the $m - \ell$ smallest right-singular vectors of M .

We first note the following fact regarding the singular values of M' given the singular vectors of M .

Fact 10.4. *Let M be an $n \times m$ matrix and let M' be an $(n + 1) \times m$ matrix obtained by adding the row X to M . Let $v_i(M)$ denote the i th right singular vector of M . Then the roots of*

$$(17) \quad 0 = \prod_{i=1}^m (\sigma_i(M)^2 - x) + \sum_{i=1}^m |\langle v_i(M), X^\dagger \rangle|^2 \prod_{j \neq i} (\sigma_j(M)^2 - x)$$

are precisely $\sigma_i(M')^2$ for $i \in [m]$.

Proof. Note that $(M')^\dagger M' = M^\dagger M + X^\dagger X$. Also, we have the rank one orthogonal decomposition

$$M^\dagger M = \sum_{i=1}^m \sigma_i(M)^2 v_i(M) v_i(M)^\dagger$$

and hence for all $x \notin \{\sigma_i(M)^2 : i \in [m]\}$ we have

$$(M^\dagger M - xI_m)^{-1} = \sum_{i=1}^m (\sigma_i(M)^2 - x)^{-1} v_i(M) v_i(M)^\dagger.$$

Next, the matrix determinant lemma gives $\det(A + uv^\dagger) = (1 + v^\dagger A^{-1}u) \det A$. Therefore by direct computation, for generic x we have

$$\det((M')^\dagger M' - xI) = (1 + X(M^\dagger M - xI_m)^{-1} X^\dagger) \det(M^\dagger M - xI_m)$$

$$= \prod_{i=1}^m (\sigma_i^2(M) - x) + \sum_{i=1}^m |\langle v_i(M), X^\dagger \rangle|^2 \prod_{j \neq i} (\sigma_j^2(M) - x).$$

The resulting equality is in fact valid for all $x \in \mathbb{C}$ by the identity theorem. Finally, the roots of $\det((M')^\dagger M' - xI) = 0$ are the squares of the singular values of M' , so the desired result follows. \square

Proof of Proposition 10.3. Note that the statement is vacuous for $\ell \geq n+1$ as the right-hand side is zero; thus it suffices to consider $\ell \leq n$. We fix unit right-singular vectors $v_1(M), \dots, v_m(M)$ such that $\|Mv_i(M)\|_2 = \sigma_i(M)$. We may write

$$M = \sum_{i=1}^n \sigma_i(M) u_i(M)^\dagger v_i(M)$$

where $u_i(M)^\dagger$ is the i th left-singular vector of M .

Via a continuity argument, it suffices to assume that $\sigma_i(M)$ are distinct for $i \leq n$ and $\langle v_i(M), X^\dagger \rangle \neq 0$ for all i . In particular, let $M^\varepsilon = M + \sum_{i=1}^n \varepsilon Z_i u_i(M)^\dagger v_i(M)$ where Z_i are uniform in $[0, 1]$ and let $X^\varepsilon = X + \varepsilon Z'$ where Z' is a standard m -dimensional Gaussian. For any sufficiently small fixed $\varepsilon > 0$, with probability 1, X^ε and M^ε satisfy $\langle v_i(M), (X^\varepsilon)^\dagger \rangle \neq 0$, M^ε has the same left- and right-singular vectors as M (up to reordering the ones which were for the same singular value), and the singular values of M^ε are distinct. Taking $\varepsilon \rightarrow 0^+$ gives the desired result.

Now define

$$F(x) = 1 + \sum_{i=1}^m \frac{|\langle v_i(M), X^\dagger \rangle|^2}{\sigma_i(M)^2 - x}.$$

Roots of $F(x)$ are $\sigma_i(M)^2$ for $i \leq \min(n+1, m)$ by Fact 10.4 and since the $\sigma_i(M)$ are distinct. $F(x)$ is increasing in $(\sigma_i(M)^2, \sigma_{i-1}(M)^2)$ (where we write $\sigma_0(M) = +\infty$) for $1 \leq i \leq n$. Finally, for $1 \leq i \leq n$ we have

$$\lim_{x \rightarrow \sigma_i(M)^2+} F(x) = -\infty \text{ and } \lim_{x \rightarrow \sigma_i(M)^2-} F(x) = +\infty.$$

Let $\eta = \sum_{i=\ell+1}^m |\langle v_i(M), X^\dagger \rangle|^2$ and note that $\eta = \|PX^\dagger\|_2^2$. For $x \in (\sigma_{\ell+1}(M)^2, \sigma_{k-1}(M)^2)$, we have

$$\begin{aligned} F(x) &\leq 1 + \frac{\sum_{i=1}^{k-1} |\langle v_i(M), X^\dagger \rangle|^2}{\sigma_{k-1}(M)^2 - x} + \sum_{i=k}^m \frac{|\langle v_i(M), X^\dagger \rangle|^2}{\sigma_i(M)^2 - x} \\ &\leq 1 + \frac{\sum_{i=1}^{k-1} |\langle v_i(M), X^\dagger \rangle|^2}{\sigma_{k-1}(M)^2 - x} + \sum_{i=k}^{\ell} \frac{|\langle v_i(M), X^\dagger \rangle|^2}{\sigma_i(M)^2 - x} + \sum_{i=\ell+1}^m \frac{|\langle v_i(M), X^\dagger \rangle|^2}{-x} \\ &= 1 + \frac{\sum_{i=1}^{k-1} |\langle v_i(M), X^\dagger \rangle|^2}{\sigma_{k-1}(M)^2 - x} + \sum_{i=k}^{\ell} \frac{|\langle v_i(M), X^\dagger \rangle|^2}{\sigma_i^2 - x} + \frac{\eta}{-x} =: G(x). \end{aligned}$$

Via direct inspection $G(x)$ has $\ell - k + 3$ zeros $\sigma_0^{*2} > \sigma_1^{*2} > \dots > \sigma_{\ell-k+2}^{*2}$. Using $F(x) \leq G(x)$ we have that $0 \leq \sigma_i^* \leq \sigma_{i+k-1}(M')$ for $1 \leq i \leq \ell - k + 2$ and $\sigma_0^* > \sigma_{k-1}(M)$. By Vieta's formula we have

$$\prod_{i=0}^{\ell-k+2} \sigma_i^{*2} = \eta \prod_{i=k-1}^{\ell} \sigma_i(M)^2.$$

Finally, note that G is increasing and starts near $-\infty$ for x close to σ_{k-1}^2 from above and

$$G(\sigma_{k-1}(M)^2 + \|X\|_2^2) \geq 1 - \frac{1}{\|X\|_2^2} \left(\sum_{i=1}^{k-1} |\langle v_i(M), X^\dagger \rangle|^2 + \sum_{i=k}^{\ell} |\langle v_i(M), X^\dagger \rangle|^2 + \eta \right) = 0.$$

Thus the additional root satisfies the bound $\sigma_0^{*2} \leq \sigma_{k-1}^2 + \|X\|_2^2$. Finally, we deduce

$$\prod_{i=k}^{\ell+1} \sigma_i(M')^2 \geq \prod_{i=1}^{\ell-k+2} \sigma_i^{*2} = \eta \sigma_0^{*-2} \prod_{i=k-1}^{\ell} \sigma_i^2 \geq \eta (\|X\|_2^2 + \sigma_{k-1}^2)^{-1} \prod_{i=k-1}^{\ell} \sigma_i^2.$$

Taking square roots completes the proof. \square

11. SINGULAR VALUE CONVERGENCE

The key output of this section is convergence of the singular value measures associated to shifted copies of B_m and B_n . As this material is a completely standard application of the method of moments, we will be brief with details.

Lemma 11.1. *Fix $z \in \mathbb{C}$ and B_m and B be as in Section 4 (treating ε as fixed). Let $\nu'_{z,n}, \nu_{z,n}$ be the empirical spectral measures of $B_m - zI$ and $B - zI$ respectively. There exist deterministic measures $\nu'_{|z|,\varepsilon}$ and $\nu_{|z|}$ such that $\nu'_{z,n} \rightsquigarrow \nu'_{|z|,\varepsilon}$ and $\nu_{z,n} \rightsquigarrow \nu_{|z|}$.*

We first prove the necessary convergence in moments.

Lemma 11.2. *Let $B_m, B, d,$ and ε be as in Section 4 (treating ε as fixed). Furthermore for a matrix M , define $M^{(1)} = M$ and $M^{(-1)} = M^\dagger$. Then for $r \leq \log \log n$, $\vec{s} = (s_1, \dots, s_r) \in \{\pm 1\}^r$ there exist constants $M(\vec{s}, \varepsilon), M'(\vec{s})$ such that*

$$\begin{aligned} \mathbb{P}\left(\left|\frac{1}{m} \operatorname{Tr} \prod_{i=1}^r B_m^{(s_i)} - M(\vec{s}, \varepsilon)\right| \geq n^{-1/2+o(1)}\right) &\lesssim n^{-\omega(1)}, \\ \mathbb{P}\left(\left|\frac{1}{n} \operatorname{Tr} \prod_{i=1}^r B^{(s_i)} - M'(\vec{s})\right| \geq n^{-1/2+o(1)}\right) &\lesssim n^{-\omega(1)}. \end{aligned}$$

Furthermore

$$\begin{aligned} \left|\mathbb{E}\left[\frac{1}{m} \sum_{i=1}^m (\sigma_i(B_m - zI)^{2r} - \sigma_i(B_m - |z|I)^{2r})\right]\right| &\lesssim_{|z|} n^{-1/2+o(1)}, \\ \left|\mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n (\sigma_i(B - zI)^{2r} - \sigma_i(B - |z|I)^{2r})\right]\right| &\lesssim_{|z|} n^{-1/2+o(1)}. \end{aligned}$$

Proof. We prove the claims only for B_m ; the analogous claim for B is rather simpler noting that B can be coupled with a $\operatorname{Ber}(d/n)$ matrix by changing at most $(\log n)^3$ many entries with probability $1 - n^{-\omega(1)}$.

Note that $\operatorname{Tr} \left(\prod_{i=1}^{\ell} B_m^{(s_i)} \right)$ can be interpreted as the number of walks of length ℓ such that the i th step is taken on the digraph associated to B if $s_i = 1$ and is taken on the digraph associated to B^\dagger if $s_i = -1$. We prove concentration conditional on the degree sequence of B_m in the configuration model. To show the different possible outcomes of the degree sequence have close means, a standard expectation computation in the configuration model and the control from Lemma 6.6 suffices.

To prove concentration in the configuration model, the pairing used to define the associated random graph can be viewed as a uniformly random permutation between the left and right stubs. Furthermore, changing two pairings in this random permutation creates or destroys at most $(\log n)^{\log \log n} = n^{o(1)}$ walks counted by the moment under the assumption that degree sequence of the digraph associated to B_m has maximum in- and out-degree bounded by $\log n$ (which we may assume, occurring with probability $1 - n^{-\omega(1)}$). The desired concentration then follows from Lemma A.3.

For the second part of the lemma for B_m , note that

$$\frac{1}{m} \sum_{i=1}^m \sigma_i(B_m - zI)^{2r} = \frac{1}{m} \operatorname{tr}(((B_m - zI)^\dagger(B_m - zI))^r).$$

One may view this as walks of length $2r$ such that the odd steps correspond to B_m with potential self-loops of weight $-z$ and the even steps correspond to backwards transversal on B_m with potential self-loops of weight $-\bar{z}$. We break up the contributions to the expectation of $\frac{1}{m} \operatorname{tr}(((B_m - zI)^\dagger(B_m - zI))^r)$ based on the precise graph-theoretic structure (within B_m), decorating special self-loops with either $-z$ or $-\bar{z}$. The underlying directed graph is connected.

In order to contribute to the leading term of the expectation, this digraph must have v vertices with $v - 1$ total directed edges (not counting special self-loops) and thus is a tree (when viewed as an undirected graph). Additionally, all self-loops must be decorated by $-z$ or $-\bar{z}$. The walk corresponds to a back-and-forth transversal of this tree which goes through each directed edge at least once in each direction, along with transversal of the decorated edges, and it returns to its original point.

Fix such a transversal and analyze the vertex w furthest from the root. We see that the walk traverses an edge to reach w , then self-loops through w an even number of times (which contributes $|z|^{2k}$ since the weights must alternate between $-z, -\bar{z}$), and then transverses back along this edge. Factoring out this contribution and proceeding downward inductively, one can observe that all dominant terms have expectations which are a function of $|z|$ (not simply z). \square

Next we require the following bound on the operator norm of a matrix in terms of the ℓ_1 norms of its rows and columns, due to Schur [28].

Lemma 11.3. *For any matrix M ,*

$$\|M\|_{\text{op}} \leq \|M\|_{1 \rightarrow 1}^{1/2} \|M^\dagger\|_{1 \rightarrow 1}^{1/2}.$$

Remark. Note that the $1 \rightarrow 1$ norm of a matrix is the maximum ℓ_1 norm of a column.

Furthermore we will require a variant of Weyl's inequality which follows immediately by Courant–Fischer theorem (for singular values).

Fact 11.4. *Fix matrices $A, B \in \mathbb{C}^{n \times n}$. We have for all $1 \leq i \leq n$ that*

$$|\sigma_i(A) - \sigma_i(B)| \leq \|A - B\|_{\text{op}}.$$

The last ingredient is control on the growth rate of the moments of the singular values, which will allow us to use Carleman's condition for moment matching of distributions.

Lemma 11.5. *Let B_m, B, d , and ε be as in Section 4. There exists a constant $C > 0$ such that with probability $1 - n^{-\omega(1)}$, for $k \leq \log \log n$ we have*

$$\sum_{i=1}^m \sigma_i(B_m - zI)^k \leq \sum_{i=1}^n \sigma_i(B - zI)^k \leq n \left(\frac{O(dk)}{\log(k+1)} + 4|z| \right)^k.$$

Proof. By Fact 10.1 and Fact 11.4 we have that $\sigma_i(B_m - zI) \leq \sigma_i(B - zI) \leq \sigma_i(B) + |z|$. Thus it suffices to consider B and consider the case when $z = 0$. Furthermore one can couple A and B so that A and B differ in at most $(\log n)^3$ entries with probability $1 - n^{-\omega(1)}$; recall that A is a matrix where all entries are independent $\text{Ber}(d/n)$.

Note that

$$\mathbb{P}(\text{Ber}(d/n) \geq t) \leq \binom{n}{t} (d/n)^t \leq \left(\frac{ed}{t} \right)^t.$$

Therefore applying Chernoff's inequality to the rows and columns of A separately implies that there are at most $(Cd/t)^t n$ vertices with degree larger t for $t \in [1, \sqrt{\log n}]$ and the maximum degree is at most $\log n$ with probability $1 - n^{-\omega(1)}$. Therefore, using Lemma 11.3 we have

$$\sum_{i=1}^n \sigma_i(A)^k \leq n + \int_{x \geq 1} kx^{k-1} (x/(Cd))^{-x/C} dx + n(\log n)^k \cdot \exp(-\Omega((\log n)^{1/2})) \leq n \left(\frac{O(dk)}{\log(k+1)} \right)^k.$$

As A and B differ in at most $(\log n)^3$ entries with appropriate probability, an analogous bound holds for B . \square

We now deduce the main output of this section.

Proof of Lemma 11.1. We will restrict attention to convergence of the singular values of $B_m - zI_m$; the proof is identical for $B_n - zI_n$. Note that

$$\sum_{i=1}^m \sigma_i(B_m - zI_m)^{2k} = \text{Tr}(((B_m - zI_m)^\dagger (B_m - zI_m))^k).$$

We use the second part of Lemma 11.2 to replace z with $|z|$ and then use expansion and the first part to deduce that there are constants $C(|z|, k)$ such that with probability $1 - n^{-\omega(1)}$ we have

$$\left| \sum_{i=1}^m \sigma_i(B_m - zI_m)^{2k} - mC(|z|, k) \right| \lesssim_{K, |z|} n^{-1/3}$$

for all $k \leq \log \log n$. Furthermore by Lemma 11.5 we have $C(|z|, k) \leq \left(\frac{O(dk)}{\log(k+1)} + 4|z| \right)^k$.

Let $\tilde{\nu}_{z,m}$ denote the uniform measure on the set $\bigcup_{1 \leq i \leq m} \{\sigma_i(B_m - zI_m), -\sigma_i(B_m - zI_m)\}$. It suffices to prove convergence in distribution of $\tilde{\nu}_{z,m}$ to a measure $\tilde{\nu}_{|z|, \varepsilon}$ by the continuous mapping theorem as $|\cdot|$ is continuous everywhere.

By a standard argument (see e.g. [8, Section 3.3.5, p. 140]), it suffices to prove that there exists a unique distribution with zero odd moments and even moments $C(|z|, k)$. Note that $\sum_{k \geq 1} C(|z|, k)^{-1/(2k)} \gtrsim_{d, |z|} \sum_{k \geq 1} (\log k)/k = \infty$ and therefore Carleman's condition applies (see e.g. [8, Theorem 3.3.25, Remark]). Also, evidently the distribution one converges to depends only on $|z|$. \square

12. NON-ATOMICITY OF SINGULAR VALUE MEASURES

The key output of this section is that the singular value measure $B_m - zI_m$ does not have an atom at zero for (Lebesgue) almost all $z \neq 0$. This will be used as the starting point for the singular value product considered in our proof.

Lemma 12.1. *Let B_m be as in Section 4. There is $c = c(d, \Xi) > 0$ such that for every $\tau \in (0, 1/2)$ and $\gamma \in (0, 1/2)$ the following holds.*

There exists a subset $\Gamma = \Gamma(d, \gamma, \Xi, \tau)$ of $\{z \in \mathbb{C} : |z| \leq \Xi\}$ such that the measure of $\{z \in \mathbb{C} : |z| \leq \Xi\} \setminus \Gamma$ is bounded by $\exp(1 - c\gamma \log(1/\tau))$ and for all $z \in \Gamma$ we have

$$\lim_{n \rightarrow \infty} \mathbb{P}(\sigma_{\lceil (1-\gamma)m \rceil}(B_m - zI_m) \leq \tau) = 0.$$

Remark. We believe that the only z for which

$$\lim_{\tau \rightarrow 0^+} \limsup_{n \rightarrow \infty} \mathbb{P}(\sigma_{\lceil (1-\gamma)m \rceil}(B_m - zI_m) \leq \tau) \neq 0$$

is $z = 0$. This is equivalent to the statement that the limiting singular value measures for $z \neq 0$ lack an atom at 0; such a technical condition is sufficient to provide us a starting point for the random walk defined in Section 4. Lemma 12.1 implies the weaker statement that the set of z which fail is Lebesgue measure zero.

We first require the following elementary lemma regarding the measure of sets with large logarithmic potential.

Lemma 12.2. *There exists an absolute constant $C > 0$ such that the following holds. Given complex numbers $(\lambda_i)_{1 \leq i \leq n}$, define*

$$S_\tau = \left\{ z \in \mathbb{C} : \prod_{i=1}^n |\lambda_i - z| \leq \exp(-\tau n) \right\}.$$

If μ denotes Lebesgue measure, then on \mathbb{C} that

$$\mu(S_\tau) \leq C \exp(-C^{-1}\tau).$$

Proof. Fix z and order λ_i such that $|\lambda_1 - z| \geq |\lambda_2 - z| \geq \dots \geq |\lambda_n - z|$. Suppose that $|\lambda_{n+1-i} - z| \geq \exp(-\tau(n/i)^{1/2})$ for each $i = 2^j$ with $1 \leq 2^j < n/2$. This implies that

$$\prod_{i=1}^n |\lambda_i - z| \geq \prod_{1 \leq 2^j < n/2} (\exp(-8^{-1}\tau n^{1/2}2^{-j/2}))^{2^j} > \exp(-\tau n).$$

Thus we must have

$$S_\tau \subseteq \bigcup_{1 \leq 2^j \leq n/2} \left\{ z \in \mathbb{C} : \#\{k : |\lambda_k - z| \leq \exp(-\tau n^{1/2}2^{-j/2})\} \geq 2^j \right\}.$$

Note that Markov's inequality gives

$$\mu\left(\left\{ z \in \mathbb{C} : \#\{k : |\lambda_k - z| \leq \exp(-\tau n^{1/2}2^{-j/2})\} \geq 2^j \right\}\right) \leq n2^{-j} \cdot \pi \exp(-2\tau n^{1/2}2^{-j/2}).$$

The result follows by summing all $1 \leq 2^j \leq n/2$. \square

We now prove Lemma 12.1.

Proof of Lemma 12.1. By adjusting c appropriately, we may assume τ is small with respect to d . Consider \mathcal{N} , a finite τ^2 -net of $\{z \in \mathbb{C} : |z| \leq \Xi\}$ with $|\mathcal{N}| \lesssim (\Xi/\tau^2)^2$.

First, consider the set \mathcal{N}' of all $z \in \mathcal{N}$ such that $\lim_{n \rightarrow \infty} \mathbb{P}(\sigma_{\lceil(1-\gamma)m\rceil}(B_m - zI_m) \leq \tau) \neq 0$. By Lemma 11.1, we find $\lim_{n \rightarrow \infty} \mathbb{P}(\sigma_{\lceil(1-\gamma/2)m\rceil}(B_m - zI_m) \leq 2\tau) = 1$. Since \mathcal{N}' is a finite set, there is some slowly decaying $\alpha(n) \rightarrow 0$ so that

$$\mathbb{P}(\sigma_{\lceil(1-\gamma/2)m\rceil}(B_m - z'I_m) \leq 2\tau \text{ for all } z' \in \mathcal{N}') \geq 1 - \alpha(n).$$

For any $z' \in \mathcal{N}'$ we can consider the τ^2 -disk around it. The union of these disks we set to be \mathcal{N}^* . We have

$$\mathbb{P}(\sigma_{\lceil(1-\gamma/2)m\rceil}(B_m - z^*I_m) \leq 3\tau \text{ for all } z^* \in \mathcal{N}^*) \geq 1 - \alpha(n)$$

by Fact 11.4 (which shows singular values shift by at most τ^2).

This event along with $\|B_m\|_{\text{HS}}^2 \leq 2dn$ (which occurs whp) implies that for $z^* \in \mathcal{N}^*$ we have

$$\prod_{j=1}^m |\lambda_j(B_m) - z| = \prod_{j=1}^m \sigma_j(B_m - zI_m) \leq (3\tau)^{\gamma m/3} (4d + 2|z|)^n \leq \exp(-\Omega(\gamma n \log(1/\tau)))$$

as long as $\gamma \log(1/\tau)$ is sufficiently large (which can be enforced by making c small enough; if this product is not large then the result is vacuous).

By Lemma 12.2, the above occurs for a set of z of measure at most $\exp(-\Omega(\gamma \log(1/\tau))) = \tau^{\Omega(\gamma)}$ if c is small enough. So, with probability at least $1 - 2\alpha(n)$ we have $\mu(\mathcal{N}^*) \leq \tau^{\Omega(\gamma)}$. Since \mathcal{N}^* is a deterministic set, we simply have $\mu(\mathcal{N}^*) \leq \tau^{\Omega(\gamma)}$.

Note that everything in the disk of radius Ξ excluding \mathcal{N}^* , call this set \mathcal{G} , is necessarily close to an element of $\mathcal{N} \setminus \mathcal{N}'$. We have

$$\lim_{n \rightarrow \infty} \mathbb{P}(\sigma_{\lceil(1-\gamma)m\rceil}(B_m - zI_m) \leq \tau) = 0$$

for $z \in \mathcal{N} \setminus \mathcal{N}'$ so a similar rounding argument to before using Fact 11.4 and $\tau^2 \leq \tau/2$ shows that

$$\lim_{n \rightarrow \infty} \mathbb{P}(\sigma_{\lceil(1-\gamma)m\rceil}(B_m - zI_m) \leq \tau/2) = 0$$

for all $z \in \mathcal{G}$. Changing the parameters γ, τ appropriately and recalling $\mu(\mathcal{N}^*) \leq \tau^{\Omega(\gamma)}$, we are done. \square

13. PROOF OF THEOREM 1.1

13.1. Reduction to tail estimate for shifted singular values. In this section we give the precise results which convert control over small singular values for shifted random matrices for (Lebesgue) almost all z into convergence of the associated spectral measure.

We now state the criterion for the convergence of spectral measures; related criteria appear in work of Tao and Vu [30]. We rely on a criterion given in work of Bordenave and Chafaï [5, Lemma 4.3, Remark 4.4]. For an $n \times n$ matrix M , we define

$$\nu_M = \frac{1}{n} \sum_{i=1}^n \delta_{\sigma_i(M)}.$$

Proposition 13.1. *Let $(M_n)_{n \geq 1}$ be a sequence of random $n \times n$ matrices. Suppose that $(\nu_z)_{z \in \mathbb{C}}$ are (non-random) probability measures on \mathbb{R}^+ such that for (Lebesgue) almost all z , for any $\varepsilon > 0$*

$$\nu_{M_n - zI} \rightsquigarrow \nu_z \text{ and } \lim_{t \rightarrow +\infty} \limsup_{n \rightarrow \infty} \mathbb{P}\left(\int_{|\log u| \geq t} |\log u| d\nu_{M_n - zI}(u) > \varepsilon\right) = 0.$$

Then there exists a probability measure μ on \mathbb{C} such that

$$\mu_{M_n} \rightsquigarrow \mu.$$

We now state a precise tail estimate which will be sufficient to deduce the main theorem.

Lemma 13.2. *Fix $d > 1$ and define B as in Section 4. Assume the associated value of ε is sufficiently small as a function of d . There exist $\delta = \delta(\varepsilon, z) > 0$ (with $\delta \rightarrow 0$ as $\varepsilon \rightarrow 0$) and $C = C(d, z) > 0$ such that for (Lebesgue) almost all $z \in \mathbb{C}$,*

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\prod_{j=0}^{\delta n} \sigma_{n-j}(B - zI) \leq \exp(-C\varepsilon n)\right) = 0.$$

We now deduce Theorem 1.1 from Lemma 13.2.

Proof of Theorem 1.1 for $d > 1$ given Lemma 13.2. Let A and B be as in Section 4. For all $z \in \mathbb{C}$, by Lemma 11.1, Lemma 5.1, and noting that permutation matrices are unitary it follows that $\nu_{A-zI} \rightsquigarrow \nu_{|z|}$.

We now verify the crucial uniform integrability condition in Proposition 13.1. Fix $\varepsilon = \mathbb{P}[\text{Pois}(d) \geq \Delta] > 0$. For almost all $z \in \mathbb{C}$, by Lemma 13.2 there is $\delta = \delta(\varepsilon, z) > 0$ such that

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{1}{n} \sum_{j=0}^{\delta n} \log(\sigma_{n-j}(B - zI)) \leq -C\varepsilon\right) = 0.$$

By the strong law of large numbers, whp $\|B\|_{\text{HS}}^2 \leq 2dn$ and thus $\sigma_{n/2}(B - zI)^2 \leq 8(d + |z|^2)$. Therefore we have

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{1}{n} \sum_{j=1}^n \log(\sigma_j(B - zI)) \mathbb{1}_{\sigma_j(B - zI) \leq \exp(-2C\varepsilon/\delta)} \leq -C\varepsilon - \delta \cdot \log(8(d + |z|^2)) \right) = 0;$$

this is the crucial estimate controlling the lower tail. For the large values of the logarithm, note that $(\log x) \mathbb{1}_{x \geq T} \leq \frac{x^2}{T}$ for $T \geq 1$. Therefore if $\|B\|_{\text{HS}}^2 \leq 2dn$ then

$$\begin{aligned} \frac{1}{n} \sum_{j=1}^n \log(\sigma_j(B - zI)) \mathbb{1}_{\sigma_j(B - zI) \geq \varepsilon^{-1}(d + |z|^2)} &\leq \frac{1}{n(\varepsilon^{-1}(d + |z|^2))} \sum_{j=1}^n \sigma_j(B - zI)^2 \\ &\leq \frac{2(\|B\|_{\text{HS}}^2 + |z|^2 n)}{n(\varepsilon^{-1}(d + |z|^2))} \leq 4\varepsilon. \end{aligned}$$

Therefore for almost all $z \in \mathbb{C}$, there exists $T = T(z, \varepsilon) > 0$ such that

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{1}{n} \sum_{j=1}^n |\log(\sigma_j(B - zI))| \mathbb{1}_{|\log(\sigma_j(B - zI))| \geq T} \geq (C + 4)\varepsilon + \delta \log(4d + |z|^2) \right) = 0.$$

By Lemma 5.1, we therefore have

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{1}{n} \sum_{j=1}^n |\log(\sigma_j(A - zI))| \mathbb{1}_{|\log(\sigma_j(A - zI))| \geq T} \geq (C + 4)\varepsilon + \delta \log(4d + |z|^2) \right) = 0.$$

Taking the countable sequence of possible ε tending to 0 given by taking $\Delta \rightarrow \infty$ and recalling $\delta(\varepsilon, z) \rightarrow 0$ as $\varepsilon \rightarrow 0$ verifies the second condition of Proposition 13.1. The result follows. \square

13.2. Proof of Lemma 13.2. We now proceed with the proof of the key Lemma 13.2. This essentially follows from piecing together the ingredients which have been developed in the paper.

Proof of Lemma 13.2. We are now finally in position to run the random walk.

Step 1: Definition of the random walk. Consider $\Xi_j = 2^j$. By Lemma 12.1 with $\gamma = \varepsilon^4$, there exists $\tau_j > 0$ such that for all but a measure 2^{-j} set of values z , we have that

$$\mathbb{P}(\sigma_{\lceil(1-\varepsilon^4)m\rceil}(B_m - zI_m) \leq \tau_j) = o(1).$$

Considering all positive integral j , it follows that for all but a measure 0 set of z there is $\tau(z) > 0$ such that

$$\mathbb{P}(\sigma_{\lceil(1-\varepsilon^4)m\rceil}(B_m - zI_m) \leq \tau(z)) = o(1).$$

Define $\delta m = \lfloor \varepsilon^4 \min((\log(1/\tau(z)))^{-1}, 1/25)m \rfloor$ and note that $\delta \leq \varepsilon^4/25$ which implies that $\delta \rightarrow 0$ as $\varepsilon \rightarrow 0$. Furthermore with probability $1 - o(1)$, we have

$$\prod_{j=\lceil(1-\varepsilon^4)m\rceil-\delta m}^{\lceil(1-\varepsilon^4)m\rceil} \sigma_j(B_m - zI_m) \geq (\tau(z))^{\delta m} \geq \min \left(1, (\tau(z))^{\varepsilon^4 m / \log(1/\tau(z))} \right) = \exp(-\varepsilon^4 m).$$

We now define the random walk. We set $X_m = m - \lceil(1 - \varepsilon^4)m\rceil \leq \varepsilon^4 m$ and therefore

$$\prod_{j=X_m}^{X_m+\delta m} \sigma_{m-j}(B_m - zI_m) \geq \exp(-\varepsilon^4 m).$$

Recall the definition of ε_r from Section 8. We iteratively define the random variable X_{t+1} when $m \leq t \leq n - \ell - 1$ by:

- If $X_t \leq \lfloor n - \ell - t \rfloor / 16$, $\|B_{t+1}\|_{\text{HS}}^2 \geq 2dn$, or $t \geq n - \ell - (\log n)^{7/4}$ define $X_{t+1} = X_t + 1$.
- Else if $\sigma_{t-X_t/2}(B_t - zI_t) \geq \varepsilon_{X_t}$, define $X_{t+1} = X_t - 1$.

- Else if

$$\text{either } \min(\deg_{B_{n-\ell}}^+(v_{t+1}, T_1), \deg_{B_{n-\ell}}^-(v_{t+1}, T_1)) \leq \sqrt{\log(1/\varepsilon)}$$

$$\text{or } \max(\deg_{B_{n-\ell}}^+(v_{t+1}), \deg_{B_{n-\ell}}^-(v_{t+1})) \geq (\log(1/\varepsilon))^2,$$

define $X_{t+1} = X_t + 1$.

- Else if

$$\|P_{X_t, B_t^\dagger - \bar{z}I_t}((B_{t+1}^* - zI_{t \times (t+1)})e_{t+1})\|_2 \geq \varepsilon_{X_t},$$

$$\|P_{X_t, B_{t+1}^* - zI_{t \times (t+1)}}((B_{t+1}^\dagger - \bar{z}I_{t+1})e_{t+1})\|_2 \geq \varepsilon_{X_t}$$

both hold, define $X_{t+1} = X_t - 1$.

- Else define $X_{t+1} = X_t + 1$.

When $n - \ell \leq t \leq n - 1$ we define X_{t+1} by:

- If $X_t > 1$ and $\sigma_t(B_t) \geq \varepsilon_1$, define $X_{t+1} = X_t - 1$.
- Else if $X_t = 1$ and $\sigma_t(B_t) \geq \varepsilon_1$, or $X_t = 0$ and

$$\|P_{1, B_{t+1}^* - zI_{t \times (t+1)}}((B_{t+1}^\dagger - \bar{z}I_{t+1})e_{t+1})\|_2 \geq \varepsilon_1,$$

define $X_{t+1} = 0$.

- Else if $X_t \geq 1$, $\sigma_t(B_t) \leq \varepsilon_1$, and

$$\|P_{1, B_t^\dagger - \bar{z}I_t}((B_{t+1}^* - zI_{t \times (t+1)})e_{t+1})\|_2 \geq \varepsilon_1,$$

$$\|P_{1, B_{t+1}^* - zI_{t \times (t+1)}}((B_{t+1}^\dagger - \bar{z}I_{t+1})e_{t+1})\|_2 \geq \varepsilon_1$$

both hold, define $X_{t+1} = X_t - 1$.

- Else define $X_{t+1} = X_t + 1$.

Step 2: Reducing to proving that $X_n = 0$ whp. Let

$$\tau_1 = 8(d(\log(2/\varepsilon))^4 + |z|^2) \text{ and } \tau_2 = 8n^4(1 + |z|^2).$$

We claim that

$$(18) \quad \prod_{j=X_{t+1}}^{X_{t+1}+\delta m} \sigma_{t+1-j}(B_{t+1} - zI_{t+1}) \geq \tau_1^{-2} \varepsilon_{X_t}^2 \prod_{j=X_t}^{X_t+\delta m} \sigma_{t-j}(B_t - zI_t)$$

for $m \leq t \leq n - \ell - 1$ and

$$(19) \quad \prod_{j=X_{t+1}}^{X_{t+1}+\delta m} \sigma_{t+1-j}(B_{t+1} - zI_{t+1}) \geq \tau_2^{-2} \varepsilon_1^2 \prod_{j=X_t}^{X_t+\delta m} \sigma_{t-j}(B_t - zI_t)$$

for $n - \ell \leq t \leq n - 1$.

Given this claim, if $X_n = 0$ then iterating shows that

$$\prod_{j=0}^{\delta m} \sigma_{n-j}(B_n - zI_n) \geq \tau_1^{-2(n-m)} \tau_2^{-2\ell} \prod_{t=m}^{n-\ell} \varepsilon_{X_t}^2 \cdot \varepsilon_1^{2\ell} \cdot \prod_{j=X_m}^{X_m+\delta m} \sigma_{m-j}(B_m - zI_m)$$

$$\geq \exp(-C'' \varepsilon n) \cdot \prod_{j=X_m}^{X_m+\delta m} \sigma_{m-j}(B_m - zI_m) \geq \exp(-2C'' \varepsilon n)$$

for a constant $C'' = C''(d, z)$. The only nontrivial term to estimate is $\prod_{t=m}^{n-\ell} \varepsilon_{X_t}^2$. The desired estimate follows noting that the definition of the walk enforces $X_t \geq \lfloor n - \ell - t \rfloor / 16 - 1$ for $m \leq t \leq n - \ell - 1$, and using $n - m \leq 2\varepsilon^3 n$. We now show why (18) and (19) hold, which will allow us to focus on proving $X_n = 0$ whp in the remainder of the proof.

When $X_{t+1} = X_t + 1$, we have that (18) and (19) hold trivially by Facts 10.1 and 10.2. The second item defining X_t in the first epoch and first item defining X_t in the second epoch are similarly handled by Facts 10.1 and 10.2 and the lower bounds on singular values that we are given in these cases.

We now explain how to deduce the claim if the fourth item of the first epoch is used to define X_{t+1} ; the remaining deductions (corresponding to the second and third items in the second epoch) are essentially identical and are omitted. Note that the norms of the additional row and column added are bounded by $(2\log(1/\varepsilon)^4 + 2|z|^2)^{1/2}$ since the in- and out-degrees of v_{t+1} are bounded by $(\log(1/\varepsilon))^2$. Furthermore note that

$$\sigma_{n/2}(B_t - zI_t)^2 \leq \frac{\|B_t - zI_t\|_{\text{HS}}^2}{n/2} \leq \frac{2\|B_t\|_{\text{HS}}^2 + 2|z|^2n}{n/2} \leq \tau_1$$

since $X_t + \delta m \leq n/3$ almost surely and $\|B_t\|_{\text{HS}}^2 \leq 2dn$.

Therefore applying Proposition 10.3 we have

$$\begin{aligned} \prod_{j=X_t}^{X_t+\delta m} \sigma_{t+1-j}(B_{t+1}^* - zI_{t \times (t+1)}) &\geq \tau_1^{-1} \varepsilon_{X_t} \prod_{j=X_t}^{X_t+\delta m} \sigma_{t-j}(B_t - zI_t), \\ \prod_{j=X_t+1}^{X_t+1+\delta m} \sigma_{t+1-j}(B_{t+1} - zI_{t+1}) &\geq \tau_1^{-1} \varepsilon_{X_t} \prod_{j=X_t}^{X_t+\delta m} \sigma_{t+1-j}(B_{t+1}^* - zI_{t \times (t+1)}). \end{aligned}$$

Note that in the first item we add a column to B_t to get B_{t+1}^* ; as we are considering right-singular values one needs Fact 10.2 to relate the left- and right-singular values (and this explains the conjugate transpose and \bar{z} in the first part of the fourth item of the first epoch). Multiplying these inequalities we derive exactly (18) in this case.

Step 3: Creating quasirandomness events and verifying they hold whp. We now define a sequence of quasirandomness events; these will ultimately be required in order to import the results of Section 8. For $m \leq t \leq n - \ell - 1$, we define

$$\mathcal{G}_{t+1} = \{B_t \in \mathcal{U}((\log n)^{3/2})\} \cap \{B_t^\dagger \in \mathcal{U}((\log n)^{3/2})\} \cap \{B_t \in \mathcal{D}\} \cap \{B_t^\dagger \in \mathcal{D}\}.$$

We define a vertex $v \in H$ to be *degree-bad* if

$$\begin{aligned} \min(\deg_{B_{n-\ell}}^+(v, T_1), \deg_{B_{n-\ell}}^-(v, T_1)) &\leq \sqrt{\log(1/\varepsilon)} \text{ or} \\ \max(\deg_{B_{n-\ell}}^+(v), \deg_{B_{n-\ell}}^-(v)) &\geq (\log(1/\varepsilon))^2. \end{aligned}$$

Furthermore for $m \leq t \leq n - \ell - 1$, let \mathcal{H}_{t+1} be the event that v_{t+1} is degree-bad and let

$$\mathcal{J}_{t+1} = \left\{ \frac{\sum_{v \in S_{n-\ell} \setminus S_t} \mathbb{1}[v \text{ is degree-bad}]}{(n-\ell) - t} \geq \varepsilon \right\}.$$

For $n - \ell \leq t \leq n - 1$, define

$$\mathcal{G}_{t+1} = \{B_t^\dagger \in \mathcal{U}((\log \log n)^2)\} \cap \{B_t^\dagger \in \mathcal{D}\}$$

and for $n - \ell \leq t \leq n - 1$, define

$$\mathcal{G}'_{t+1} = \{B_{t+1}^* \in \mathcal{U}((\log \log n)^2)\} \cap \{B_{t+1}^* \in \mathcal{D}\} \cap \{B_{t+1}^* \in \mathcal{U}^*\}.$$

We now stitch together various claims regarding these quasi-randomness events.

Claim 13.3. *For all $t \leq n - \ell - (\log n)^{7/4}$, we have*

$$\mathbb{P}(\mathcal{J}_{t+1}) \leq n^{-\omega(1)} \quad \text{and} \quad \mathbb{P}(\mathcal{H}_{t+1} | \mathcal{J}_{t+1} \cup \mathcal{F}_{t+1}) \leq \varepsilon.$$

Proof. By Lemma 6.7 at most a $2\varepsilon^2$ fraction of vertices in H are degree-bad. Since we add vertices in a random order back to form S_j for $m \leq j \leq n - \ell$ the first result follows by Chernoff (see e.g. Lemma A.1) for the hypergeometric distribution. The second follows by noting that \mathcal{J}_{t+1} guarantees that at most an ε fraction of remaining vertices are degree-bad. \square

We next have the following consequence of Lemma 8.1.

Claim 13.4. *There is $C = C(d) > 0$ such that for all $t \leq n - \ell - (\log n)^{7/4}$, we have*

$$\mathbb{P}(X_{t+1} \leq X_t - 1 | \mathcal{J}_{t+1}, \mathcal{F}_{t+1}, \mathcal{G}_{t+1}, X_t > \lfloor n - \ell - t \rfloor / 16) \geq 1 - C(\log(1/\varepsilon))^{-1/4}.$$

Proof. By Claim 13.3, we have

$$\mathbb{P}(\mathcal{H}_{t+1} | \mathcal{J}_{t+1}, \mathcal{F}_{t+1}) \leq \varepsilon.$$

Therefore it suffices to prove that

$$\mathbb{P}(X_{t+1} \leq X_t - 1 | \mathcal{H}_{t+1}^c, \mathcal{F}'_{t+1}, \mathcal{G}_{t+1}, X_t > \lfloor n - \ell - t \rfloor / 16) \leq C(\log(1/\varepsilon))^{-1/4} / 2.$$

Note that \mathcal{G}_{t+1} (and in particular \mathcal{D}) and \mathcal{H}_{t+1}^c guarantees that $\|B_{t+1}\|_{\text{HS}}^2 \leq 2dn$. If we have $\sigma_{t-X_t/2}(B_t) \geq \varepsilon_{X_t}$, we instantaneously have $X_{t+1} = X_t - 1$ as desired. Otherwise applying Lemma 8.1 implies that the fourth item defining X_{t+1} in the first epoch holds with the desired probability bound. Note that $t \leq n - \ell - (\log n)^{7/4}$ implies that $X_t \geq (\log n)^{7/4} / 32$ due to the definition of the walk and therefore the necessary dimension lower bound to apply Lemma 8.1 holds. \square

In an analogous manner we have the following drift statement for the second epoch; the proof is analogous except we invoke Lemma 8.2 and keep track of events such as \mathcal{U}^* , hence the inclusion of \mathcal{G}'_{t+1} .

Claim 13.5. *There is $C = C(d) > 0$ such that if $X_{n-\ell} \leq 4(\log n)^{7/4}$ then for all $n - \ell \leq t \leq n - 1$, we have*

$$\mathbb{P}(\{X_{t+1} \leq X_t - 1 + \mathbb{1}_{X_t=0}\} \text{ or } \mathcal{G}_{t+1}^c \text{ or } \mathcal{G}'_{t+1} | \mathcal{F}_{t+1}) \geq 1 - C(\log n)^{-1/4}.$$

Step 4: Proving that $X_n = 0$ whp. For $m \leq t \leq n - \ell$, we define

$$Y_t = X_t \cdot \mathbb{1}\left(\bigcap_{j \leq t} \mathcal{G}_j \cap \bigcap_{j \leq t} \mathcal{J}_j\right).$$

By Claim 13.4, we have

$$\mathbb{P}(Y_{t+1} \leq Y_t - 1 + \mathbb{1}_{Y_t=0} | \mathcal{F}_{t+1}) \geq 1 - C(\log(1/\varepsilon))^{-1/4}$$

if $t \leq n - \ell - (\log n)^{7/4}$ and $Y_t > \lfloor n - \ell - t \rfloor / 16$.

Let $Z_t = Y_t$ for $t \leq n - \ell - (\log n)^{7/4}$ and $Z_{t+1} = Z_t - 1 + \mathbb{1}_{Z_t=0}$ for $n - \ell - (\log n)^{7/4} \leq t \leq n - \ell - 1$. By the random walk Lemma 9.1 we have

$$\mathbb{P}(Z_{n-\ell} \geq (\log n)^{7/4}) \leq n^{-\omega(1)}.$$

This implies that

$$\mathbb{P}(Y_{n-\ell-(\log n)^{7/4}} \leq 3(\log n)^{7/4}) \leq n^{-\omega(1)}.$$

Note here that when we apply Lemma 9.1, we require that $C(\log(1/\varepsilon))^{-1/4}$ is smaller than an absolute constant which requires ε to be small as a function of d only.

By Lemmas 6.10 and 6.11, and by Claim 13.3, we have

$$\mathbb{P}(Y_{n-\ell-(\log n)^{7/4}} \neq X_{n-\ell-(\log n)^{7/4}}) \leq n^{-1+o(1)}.$$

Therefore

$$\mathbb{P}(X_{n-\ell} \leq 4(\log n)^{7/4}) \geq 1 - n^{-1+o(1)}.$$

Next we define for $n - \ell \leq t \leq n$

$$Y'_t = X_t \cdot \mathbb{1} \left(\bigcap_{n-\ell+1 \leq j \leq t} \mathcal{G}_j \cap \bigcap_{n-\ell+1 \leq j \leq t} \mathcal{G}'_j \right).$$

By Claim 13.5 for $n - \ell \leq t \leq n - 1$, we have

$$\mathbb{P}(Y'_{t+1} \leq Y'_t - 1 + \mathbb{1}_{Y_t=0} | \mathcal{F}_t) \geq 1 - C(\log n)^{-1/4}.$$

Given $Y'_{n-\ell} \leq 4(\log n)^{7/4}$ at the start, which occurs with probability at least $1 - n^{-1+o(1)}$, and recalling $\ell = \lfloor (\log n)^2 \rfloor$, we have by Lemma 9.1 and Markov's inequality that

$$\mathbb{P}(Y'_n > 0) \lesssim_d (\log n)^{-1/8}.$$

By Lemmas 6.10, 6.12, and 6.13 we find

$$\mathbb{P}(Y'_n \neq X_n) \leq (\log n)^{-\omega(1)}$$

and therefore

$$\mathbb{P}(X_n \neq 0) \lesssim_d (\log n)^{-1/8}.$$

This (finally) completes the proof. \square

Remark 13.6. The logarithmic probability guarantee is immediate from the proof given as well as noting that Lemma 12.1 can be made sufficiently quantitative. (This is essentially immediate since Lemma 11.2 holds with sufficiently strong probability.)

For an improved bound on the least singular value, notice that in the above argument we directly considered the product of last δn singular values. Proving that $X_{n-\sqrt{n}} \leq \sqrt{n}/4$ whp, one can see that the $(\sqrt{n}/4)$ th smallest singular value of $B_{n-\sqrt{n}}$ is at least $\exp(-O(\sqrt{n}))$ with high probability. Iteratively applying Proposition 10.3 with $\ell = k - 1$ allows one to push the index of the unique singular value under consideration until it becomes the least singular value, at the cost of a factor $n^{-O(1)}$ each time. This gives the quality of bound mentioned in the remark following Theorem 1.2. It remains an interesting question whether the random walk approach used in this paper can prove a least singular value bound with probability quality $n^{-\Omega(1)}$ or with singular value size $\exp(-(\log n)^{O(1)})$. Such singular value estimates may prove useful when considering finer local laws.

ACKNOWLEDGEMENTS

This project was carried out, in part, while MS was visiting Cambridge in the 2022-2023 academic year. MS would like to thank the Cambridge combinatorics community for creating a pleasant working environment during this time. We would also like to thank Marcus Michelen for some comments on a draft of this paper.

APPENDIX A. VARIOUS PRELIMINARIES

A.1. Directed graphs, bipartite graphs, and matrices. Throughout this proof, we will pass freely between the notions of matrices, bipartite graphs, and digraphs. Given an $m \times n$ matrix M with entries in $\{0, 1\}$, we may identify this bipartite graph with n vertices on the left and m vertices on the right and an edge between $j \in [n]$ and $i \in [m]$ if and only if $M_{ij} = 1$. If $m \leq n$, we may identify M with a digraph by adding $n - m$ empty right vertices, directing all edges from left to right, and gluing corresponding vertices in the obvious manner. We write $\deg_M^+(v, S)$ for the number of out-neighbors v has in S (including a self-loop), and similar for $\deg_M^-(v, S)$; we drop S to refer to the total out- or in-degree. The *degree sequence* of M (or equivalently, of the corresponding bipartite graph or digraph) is $(\mathbf{d}, \mathbf{d}')$ where $\mathbf{d} = (\deg_M^+(v))_{v \in [n]}$ and $\mathbf{d}' = (\deg_M^-(v))_{v \in [m]}$.

For the entire paper we will concern ourselves with matrices such that $n \in \{m, m + 1\}$. Let $I_{m \times n}$ denote the matrix such that $(I_{m \times n})_{ij} = 1$ if $i = j \leq \min(m, n)$ and 0 otherwise; this aligns with the standard definition of the identity matrix for square matrices. The matrix $M - zI_{m \times n}$ can be identified as a weighted bipartite graph with n vertices on the left and m vertices on the right (with possible weights of $0, 1, -z, 1 - z$).

A.2. Concentration inequalities. We state a Chernoff bound for binomial and hypergeometric distributions (see for example [16, Theorems 2.1, 2.10]).

Lemma A.1 (Chernoff bound). *Let X be either:*

- *a sum of independent random variables, each of which take values in $\{0, 1\}$, or*
- *hypergeometrically distributed (with any parameters).*

Then for any $\delta > 0$ we have

$$\mathbb{P}[X \leq (1 - \delta)\mathbb{E}X] \leq \exp(-\delta^2\mathbb{E}X/2), \quad \mathbb{P}[X \geq (1 + \delta)\mathbb{E}X] \leq \exp(-\delta^2\mathbb{E}X/(2 + \delta)).$$

We will require a version of the classical Bernstein inequality; this appears as [33, Theorem 2.8.1].

Theorem A.2. *For a random variable X define the ψ_1 -norm*

$$\|X\|_{\psi_1} = \inf\{t > 0: \mathbb{E}[\exp(|X|/t)] \leq 2\}.$$

There is an absolute constant $c > 0$ such that the following holds. If X_1, \dots, X_N are independent random variables then

$$\mathbb{P}\left(\left|\sum_{i=1}^N X_i\right| \geq t\right) \leq 2 \exp\left(-c \min\left(\frac{t^2}{\sum_{i=1}^N \|X_i\|_{\psi_1}^2}, \frac{t}{\max_i \|X_i\|_{\psi_1}}\right)\right)$$

for all $t \geq 0$.

We will also require a standard concentration inequality for Lipschitz functions with respect to the symmetric group (and with respect to injections); this appears as [10, Lemma 3.3].

Lemma A.3. *Let $m \in \mathbb{N}$, let S be a finite set with $|S| \geq m$, let \mathcal{F} be the set of functions $\{1, \dots, m\} \rightarrow S$ and let $\mathcal{I} \subseteq \mathcal{F}$ be the set of injections $\{1, \dots, m\} \rightarrow S$. Consider a function $f: \mathcal{F} \rightarrow \mathbb{R}$ with the property $|f(\pi) - f(\pi')| \leq \sum_{i=1}^m c_i \mathbb{1}_{\pi(i) \neq \pi'(i)}$. Let $\pi \in \mathcal{I}$ be a uniformly random injection. Then for $t \geq 0$,*

$$\mathbb{P}(|f(\pi) - \mathbb{E}f(\pi)| \geq t) \leq 2 \exp\left(-\frac{t^2}{8 \sum_{i=1}^m c_i^2}\right).$$

Finally we will require the Azuma–Hoeffding inequality (see [16, Theorem 2.25]).

Lemma A.4 (Azuma–Hoeffding inequality). *Let X_0, \dots, X_n form a martingale sequence such that $|X_k - X_{k-1}| \leq c_k$ almost surely. Then*

$$\mathbb{P}(|X_0 - X_n| \geq t) \leq 2 \exp\left(-\frac{t^2}{2 \sum_{k=1}^n c_k^2}\right)$$

A.3. Configuration model. We will also require the definition of the configuration model for a bipartite graph.

Definition A.5. *Consider a pair of degree sequences $\mathbf{d} = (d_1, \dots, d_n)$ and $\mathbf{d}' = (d'_1, \dots, d'_m)$ such that $\sum d_i = \sum d'_i$. Consider a set of $r = \sum d_i + \sum d'_i$ “stubs”, n left buckets, and m right buckets. Assign d_i stubs to the i th left bucket and d'_i stubs to the i th right bucket. A configuration is a perfect matching between the $r/2$ stubs assigned to the left buckets and $r/2$ stubs assigned to right buckets. Given a configuration, contracting each of the buckets to a single vertex gives*

rise to a bipartite multigraph with degree sequence d_1, \dots, d_n in the left and d'_1, \dots, d'_m in the right.

A random bipartite graph G drawn from the configuration model with degree sequences $\mathbb{G}(\mathbf{d}, \mathbf{d}')$ is the bipartite multigraph arising from choosing the perfect matching between the left and right stubs uniformly at random.

Note that we may implicitly identify vertices on the left and right by identifying the i th vertex on the left and i th vertex on the right to obtain a digraph as in Appendix A.1. We have the following fact regarding the configuration model; the first is obvious by construction while the second is an immediate consequence of the results of Janson [15] (although many earlier results e.g. [3, 4, 21] would suffice).

Lemma A.6. *Sample $G \sim \mathbb{G}(\mathbf{d}, \mathbf{d}')$.*

- *Conditioned on being simple, G is a uniformly random bipartite graph with degree sequence \mathbf{d} on the left and \mathbf{d}' on the right.*
- *If d_i, d'_i are positive integers and $\sum d_i^2 + \sum d'_i{}^2 \leq C(n+m)$ and $n/C \leq m \leq Cn$, we have that G is simple with probability $\Omega_C(1)$.*

REFERENCES

- [1] Z. D. Bai, *Circular law*, Ann. Probab. **25** (1997), 494–529.
- [2] Anirban Basak and Mark Rudelson, *The circular law for sparse non-Hermitian matrices*, Ann. Probab. **47** (2019), 2359–2416.
- [3] Edward A. Bender and E. Rodney Canfield, *The asymptotic number of labeled graphs with given degree sequences*, J. Combinatorial Theory Ser. A **24** (1978), 296–307.
- [4] Béla Bollobás, *A probabilistic proof of an asymptotic formula for the number of labelled regular graphs*, European J. Combin. **1** (1980), 311–316.
- [5] Charles Bordenave and Djilil Chafaï, *Around the circular law*, Probab. Surv. **9** (2012), 1–89.
- [6] Simon Coste, *Sparse matrices: convergence of the characteristic polynomial seen from infinity*, Electron. J. Probab. **28** (2023), Paper No. 8, 40.
- [7] Kevin P. Costello, Terence Tao, and Van Vu, *Random symmetric matrices are almost surely nonsingular*, Duke Math. J. **135** (2006), 395–413.
- [8] Rick Durrett, *Probability—theory and examples*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 49, Cambridge University Press, Cambridge, 2019, Fifth edition.
- [9] Alan Edelman, *The probability that a random real Gaussian matrix has k real eigenvalues, related distributions, and the circular law*, J. Multivariate Anal. **60** (1997), 203–232.
- [10] Asaf Ferber, Matthew Kwan, Ashwin Sah, and Mehtaab Sawhney, *Singularity of the k -core of a random graph*, Duke Mathematical Journal **1** (2023), 1–40.
- [11] Asaf Ferber, Matthew Kwan, and Lisa Saueremann, *Singularity of sparse random matrices: simple proofs*, Combin. Probab. Comput. **31** (2022), 21–28.
- [12] V. L. Girko, *The circular law*, Teor. Veroyatnost. i Primenen. **29** (1984), 669–679.
- [13] Margalit Glasgow, Matthew Kwan, Ashwin Sah, and Mehtaab Sawhney, *The exact rank of sparse random graphs*, arXiv:2303.05435.
- [14] Friedrich Götze and Alexander Tikhomirov, *The circular law for random matrices*, Ann. Probab. **38** (2010), 1444–1491.
- [15] Svante Janson, *The probability that a random multigraph is simple. II*, J. Appl. Probab. **51A** (2014), 123–137.
- [16] Svante Janson, Tomasz Łuczak, and Andrzej Ruciński, *Random graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000.
- [17] Richard M. Karp, *The transitive closure of a random digraph*, Random Structures Algorithms **1** (1990), 73–93.
- [18] A. Kolmogorov, *Sur les propriétés des fonctions de concentrations de M. P. Lévy*, Ann. Inst. H. Poincaré **16** (1958), 27–34.
- [19] Alexander E. Litvak, Anna Lytova, Konstantin Tikhomirov, Nicole Tomczak-Jaegermann, and Pierre Youssef, *Circular law for sparse random regular digraphs*, J. Eur. Math. Soc. (JEMS) **23** (2021), 467–501.
- [20] Tomasz Łuczak, *The phase transition in the evolution of random digraphs*, J. Graph Theory **14** (1990), 217–223.

- [21] Brendan D. McKay and Nicholas C. Wormald, *Asymptotic enumeration by degree sequence of graphs with degrees $o(n^{1/2})$* , *Combinatorica* **11** (1991), 369–382.
- [22] M. L. Mehta, *Random matrices and the statistical theory of energy levels*, Academic Press, New York-London, 1967.
- [23] Guangming Pan and Wang Zhou, *Circular law, extreme singular values and potential theory*, *J. Multivariate Anal.* **101** (2010), 645–656.
- [24] Boris A. Rogozin, *On the increase of dispersion of sums of independent random variables*, *Teor. Veroyatnost. i Primenen* **6** (1961), 106–108.
- [25] Mark Rudelson and Konstantin Tikhomirov, *The sparse circular law under minimal assumptions*, *Geom. Funct. Anal.* **29** (2019), 561–637.
- [26] Mark Rudelson and Roman Vershynin, *The Littlewood-Offord problem and invertibility of random matrices*, *Adv. Math.* **218** (2008), 600–633.
- [27] Ashwin Sah, Julian Sahasrabudhe, and Mehtaab Sawhney, *The sparse circular law, revisited*, Forthcoming.
- [28] J. Schur, *Bemerkungen zur Theorie der beschränkten Bilinearformen mit unendlich vielen Veränderlichen*, *J. Reine Angew. Math.* **140** (1911), 1–28.
- [29] Terence Tao and Van Vu, *Random matrices: the circular law*, *Commun. Contemp. Math.* **10** (2008), 261–307.
- [30] Terence Tao and Van Vu, *Random matrices: universality of ESDs and the circular law*, *Ann. Probab.* **38** (2010), 2023–2065, With an appendix by Manjunath Krishnapur.
- [31] Terence Tao and Van H. Vu, *Inverse Littlewood-Offord theorems and the condition number of random discrete matrices*, *Ann. of Math. (2)* **169** (2009), 595–632.
- [32] Konstantin Tikhomirov, *Quantitative invertibility of non-hermitian random matrices*, arXiv:2206.00601.
- [33] R. Vershynin, *High-dimensional probability*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 47, Cambridge University Press, Cambridge, 2018, An introduction with applications in data science, With a foreword by Sara van de Geer.
- [34] Eugene P. Wigner, *On the distribution of the roots of certain symmetric matrices*, *Ann. of Math. (2)* **67** (1958), 325–327.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY. DEPARTMENT OF MATHEMATICS.

Email address: `asah@mit.edu`

UNIVERSITY OF CAMBRIDGE. DEPARTMENT OF PURE MATHEMATICS AND MATHEMATICAL STATISTICS.

Email address: `jdrs2@cam.ac.uk`

MASSACHUSETTS INSTITUTE OF TECHNOLOGY. DEPARTMENT OF MATHEMATICS.

Email address: `msawhney@mit.edu`