SHARP TRANSITION OF THE INVERTIBILITY OF THE ADJACENCY MATRICES OF SPARSE RANDOM GRAPHS

ANIRBAN BASAK* AND MARK RUDELSON†

Abstract. We consider three different models of sparse random graphs: undirected and directed Erdős-Rényi graphs, and random bipartite graph with an equal number of left and right vertices. For such graphs we show that if the edge connectivity probability $p \in (0,1)$ satisfies $np \geq \log n + k(n)$ with $k(n) \to \infty$ as $n \to \infty$, then the adjacency matrix is invertible with probability approaching one (here n is the number of vertices in the two former cases and the number of left and right vertices in the latter case). If $np \leq \log n - k(n)$ then these matrices are invertible with probability approaching zero, as $n \to \infty$. In the intermediate region, when $np = \log n + k(n)$, for a bounded sequence $k(n) \in \mathbb{R}$, the event Ω_0 that the adjacency matrix has a zero row or a column and its complement both have non-vanishing probability. For such choices of p our results show that conditioned on the event Ω_0^c the matrices are again invertible with probability tending to one. This shows that the primary reason for the non-invertibility of such matrices is the existence of a zero row or a column.

The bounds on the probability of the invertibility of these matrices are a consequence of quantitative lower bounds on their smallest singular values. Combining this with an upper bound on the largest singular value of the centered version of these matrices we show that the (modified) condition number is $O(n^{1+o(1)})$ on the event that there is no zero row or column, with large probability. This matches with von Neumann's prediction about the condition number of random matrices up to a factor of $n^{o(1)}$, for the entire range of p.

Contents

Date: September 24, 2018.

2010 *Mathematics Subject Classification.* 46B09, 60B20.

Key words and phrases. Random matrices, sparse matrices, Erdős-Rényi graph, invertibility, smallest singular value, condition number.

*Research partially supported by grant 147/15 from the Israel Science Foundation and ICTS–Infosys Excellence Grant.

† Part of this research was performed while this author visited Weizmann Institute of Science in Rehovot, Israel, where he held Rosy and Max Varon Professorship. He is grateful to Weizmann Institute for its hospitality and for creating an excellent work environment. The research was supported in part by the NSF grant DMS 1464514 and by a fellowship from the Simons Foundation.

1. Introduction

For an $n \times n$ real matrix A_n its *singular values* $s_k(A_n)$, $k \in [n] := \{1, 2, \ldots, n\}$ are the eigenvalues of $|A_n| := \sqrt{A_n^* A_n}$ arranged in a non-increasing order. The maximum and the minimum singular values, often of particular interest, can also be defined as

$$s_{\max}(A_n) := s_1(A_n) := \sup_{x \in S^{n-1}} \|A_n x\|_2, \qquad s_{\min}(A_n) := s_n(A_n) := \inf_{x \in S^{n-1}} \|A_n x\|_2,$$

where $S^{n-1} := \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$ and $\|\cdot\|_2$ denotes the Euclidean norm of a vector. Further, let

$$\Omega_0 := \{A_n \text{ has either a zero row or a zero column}\}. \tag{1.1}$$

Obviously, $s_{\min}(A_n) = 0$ for any $A_n \in \Omega_0$. For matrices with i.i.d. (independent and identically distributed) Bernoulli entries we establish the following sharp transition of invertibility.

Theorem 1.1. *Let A_n be an $n \times n$ matrix with i.i.d.* $\text{Ber}(p)$ *entries. That is, for $i, j \in [n]$*

$$\mathbb{P}(a_{i,j} = 1) = p, \qquad \mathbb{P}(a_{i,j} = 0) = 1 - p,$$

where $a_{i,j}$ is the (i,j)-th entry of A_n.

(i) *If $np \ge \log(1/p)$ then there exist absolute constants $0 < c_{1.1}, \widetilde{C}_{1.1}, C_{1.1} < \infty$ such that for any $\varepsilon > 0$, we have*

$$\mathbb{P}\left(\left\{s_{\min}(A_n) \le c_{1.1}\varepsilon \exp\left(-\widetilde{C}_{1.1}\frac{\log(1/p)}{\log(np)}\right)\sqrt{\frac{p}{n}}\right\} \cap \Omega_0^c\right) \le \varepsilon^{1/5} + \frac{C_{1.1}}{\sqrt[4]{np}}. \tag{1.2}$$

(ii) *If $np \le \log(1/p)$ then there exists an absolute constant $\bar{C}_{1.1}$ such that*

$$\mathbb{P}(\Omega_0) \ge 1 - \frac{\bar{C}_{1.1}}{\log n}.$$

Theorem 1.1 is a consequence of Theorem 1.7 which is proved under a more general set-up including, in particular, symmetric Bernoulli matrices. Throughout the paper $p = p_n$ may depend on n. For ease of writing we suppress this dependence.

To understand the implication of Theorem 1.1 we see that studying the invertibility property of any given matrix amounts to understanding the following three different aspects of it. Probably the easiest one is to find the probability that a random matrix is singular. For any given random matrix A_n, it means to find a bound on $\mathfrak{p}_n$, where

$$\mathfrak{p}_n := \mathfrak{p}(A_n) := \mathbb{P}(A_n \text{ is non-invertible}) = \mathbb{P}(\det(A_n) = 0) = \mathbb{P}(s_{\min}(A_n) = 0).$$

If the entries of A_n have densities with respect to the Lebesgue measure then $\mathfrak{p}_n = 0$. However, for matrices with discrete entries, the problem of evaluating the singularity probability $\mathfrak{p}_n$ makes sense.

The second question regarding the invertibility is more of a quantitative nature. There one is interested in finding the distance between A_n and the set of all singular matrices. As

$$s_{\min}(A_n) = \inf\{\|A_n - B\| : \det(B) = 0\},$$

where $\|A_n - B\|$ denotes the operator norm of the matrix $A_n - B$, a lower bound on $s_{\min}(A_n)$ yields a quantitative measure of invertibility.

The third direction, probably the most difficult, is to find the main reason for non-invertibility of a given random matrix. To elaborate on this let us consider the following well known conjecture:

Conjecture 1.2. Let A_n be a $n \times n$ matrix with i.i.d. Rademacher random variables (± 1 with equal probability). Then

$$\mathfrak{p}_n = \left(\frac{1}{2} + o(1)\right)^n,$$

where we recall that the notation $a_n = o(b_n)$ means $\lim_{n\to\infty} a_n/b_n = 0$.

It can be noted that the probability that there exist either two columns or two rows of A_n that are identical up to a change in sign is at least

$$cn^2 \left(\frac{1}{2}\right)^n = \left(\frac{1}{2} + o(1)\right)^n,$$

for some absolute constant $c > 0$. This matches with the conjectured leading order of the singularity probability. Thus the main reason for the singularity for a matrix with i.i.d. Rademacher entries is conjectured to be the existence of two identical columns or rows, up to a reversal of sign.

Theorem 1.1 addresses all three different aspects of invertibility for sparse random matrices. As it yields a lower bound on the smallest singular value it readily gives a quantitative estimate on the invertibility of matrices with i.i.d. Bernoulli entries. Setting $\varepsilon = 0$ in Theorem 1.1(i) we obtain a bound on the singularity probability. For a sharper bound on the singularity probability we refer the reader to Remark 4.5.

Probably, the most important feature of Theorem 1.1 is that it identifies the existence of a zero row or a column as the primary reason for non-invertibility. To see this, let us denote

$$\Omega_{\text{col}} := \{A_n \text{ has a zero column}\} \quad \text{and} \quad \Omega_{\text{row}} := \{A_n \text{ has a zero row}\}. \tag{1.3}$$

As the entries of A_n are i.i.d. $\text{Ber}(p)$ it is immediate that

$$\mathbb{P}(\Omega_{\text{col}}) = \mathbb{P}(\Omega_{\text{row}}) = 1 - (1 - (1-p)^n)^n. \tag{1.4}$$

This sows that if $np \ge \log n + k(n)$ then

$$\mathbb{P}(\Omega_{\text{col}}) \to 0, \quad \text{as } n \to \infty,$$

whereas for $np \le \log n - k(n)$ one has

$$\mathbb{P}(\Omega_{\text{col}}) \to 1, \quad \text{as } n \to \infty,$$

where $k(n) \to \infty$ as $n \to \infty$. As $np = \log(1/p)$ implies that $np = \log n - \delta_n \log\log n$, for some $\delta_n \sim 1$, from Theorem 1.1 we therefore deduce the following corollary.

Corollary 1.3. *Let A_n be a matrix with i.i.d.* $\text{Ber}(p)$ *entries. Then we have the following:*

(a) *If $np = \log n + k(n)$, where $\{k(n)\}_{n\in\mathbb{N}}$ is such that $k(n) \to \infty$ as $n \to \infty$, then*

$$\mathbb{P}(\Omega_0^c) \to 1 \quad \textit{and} \quad \mathbb{P}(A_n \textit{ is invertible}) \to 1 \quad \textit{as } n \to \infty.$$

(b) *If $np = \log n - k(n)$ then*

$$\mathbb{P}(\Omega_0^c) \to 0 \quad \textit{and} \quad \mathbb{P}(A_n \textit{ is invertible}) \to 0 \quad \textit{as } n \to \infty.$$

(c) *Moreover*

$$\mathbb{P}(A_n \textit{ is invertible} \mid \Omega_0^c) \to 1 \quad \textit{whenever } \mathbb{P}(\Omega_0^c) \cdot (\log n)^{1/4} \to \infty \quad \textit{as } n \to \infty.$$

Corollary 1.3(a)-(b) shows that the invertibility of a matrix with i.i.d. Bernoulli entries undergoes a sharp transition essentially at $p = \frac{\log n}{n}$. On the event Ω_0 the matrix is trivially singular. The importance of Corollary 1.3(c) lies in the fact that it shows that even when Ω_0 has a non-trivial probability, on the event Ω_0^c there is an exceptional set of negligible probability outside which the matrix is again invertible with large probability. This clearly indicates that the main reason for the non-invertibility of a matrix with i.i.d. Bernoulli entries is the existence of a zero row or a column. Moreover, the same phenomenon occurs for other classes of sparse random matrices with some dependence between the entries symmetric with respect to the diagonal (see Theorem 1.7 for a precise assumption). To the best of our knowledge this is the first instance where the primary reason of the non-invertibility for some class of random matrices has been rigorously established.

Understanding the singularity probability and the analysis of extremal singular values of random matrices have applications in compressed sensing, geometric functional analysis, theoretical computer science, and many other fields of science. Moreover, to find the limiting spectral distribution of any non-Hermitian random matrix ensemble one essentially needs to go via Girko's Hermitization technique which requires a quantitative lower bound on the smallest singular value. This has spurred a renewed interest in studying the smallest singular value. Using the bounds on the smallest singular value and implementing Girko's Hermitization technique there have been numerous works in this direction over the last fifteen years. We refer the reader to [3, 4, 6, 7, 12, 30, 35, 38, 44, 51], the survey articles [8, 40], and the references therein.

The study of the smallest singular value of a random matrix dates back to 1940's when von Neumann and his collaborators used random matrices to test their algorithm for the inversion of large matrices. They speculated that

$$s_{\min}(A_n) \sim n^{-1/2}, \qquad s_{\max}(A_n) \sim n^{1/2}, \qquad \text{with high probability} \tag{1.5}$$

(see [46, pp. 14, 477, 555] and [47, Section 7.8]), where the notation $a_n \sim b_n$ implies that $0 < \liminf_{n\to\infty} a_n/b_n \le \limsup_{n\to\infty} a_n/b_n < \infty$. Therefore, the condition number, which often serves as a measure of stability in matrix algorithms in numerical linear algebra,

$$\sigma(A_n) := \frac{s_{\max}(A_n)}{s_{\min}(A_n)} \sim n, \qquad \text{with high probability.} \tag{1.6}$$

A more precise formulation of this conjecture can be found in [39].

For matrices with i.i.d. standard normal entries Edelman [16] showed that

$$\mathbb{P}(s_{\min}(A_n) \le \varepsilon n^{-1/2}) \sim \varepsilon, \quad \text{for } \varepsilon \in (0,1). \tag{1.7}$$

On the other hand Slepian's inequality and standard Gaussian concentration inequality for Lipschitz functions yield that

$$\mathbb{P}(s_{\max}(A_n) \ge 2n^{1/2} + t) \le \exp(-t^2/2). \tag{1.8}$$

Therefore combining (1.7)-(1.8) one deduces (1.5)-(1.6) for Gaussian matrices. The prediction for general matrices remained open for a very long time.

For example, even finding bounds on the singularity probability for matrices with i.i.d. Rademacher entries turned out to be a non-trivial task. Back in the 60's Komlós [22, 23] obtained a non-trivial estimate on the singularity probability. His proof was later refined in [24] to show that the singularity probability

$$\mathfrak{p}_n = O(n^{-1/2}),$$

where the notation $a_n = O(b_n)$ implies that $\limsup_{n\to\infty} a_n/b_n < \infty$. It took more than twenty-five years to prove that $\mathfrak{p}_n$ decays exponentially fast in n. In [21] it was shown that $\mathfrak{p}_n \le c^n$ for some

$c \in (0,1)$. Subsequently the constant c was improved in [41, 42] and currently the best known bound is $c = 1/\sqrt{2} + o(1)$ (see [10]).

A quantitative estimate of the invertibility of such matrices follows from [34] where a lower bound on the smallest singular value of matrices with i.i.d. centered sub-Gaussian entries were derived. The assumption on the entries were later relaxed in [36] which enabled them to consider matrices with entries having a finite fourth moment. Under this assumption it was shown that for any $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\mathbb{P}(s_{\min}(A_n) \le \varepsilon n^{-1/2}) \le \frac{\delta}{2}. \tag{1.9}$$

Furthermore, from [26] it follows that for any $\delta > 0$ there exists K large enough so that

$$\mathbb{P}(s_{\max}(A_n) \ge K n^{1/2}) \le \frac{\delta}{2}.$$

Hence, one finds that for such matrices

$$\mathbb{P}(\sigma(A_n) \ge K\varepsilon^{-1} n) \le \delta. \tag{1.10}$$

This establishes von Neumann's prediction for the condition number for general matrices with i.i.d. centered entries having finite fourth moments. If the entries are sub-Gaussian the results of [36] further show that the probability bounds in (1.9) and (1.10) can be improved to $C\varepsilon + c^n$ for some large constant C and $c \in (0,1)$ that depend polynomially on the sub-Gaussian norm of the entries. We emphasize that one cannot obtain a probability estimate similar to (1.7), as Rademacher random variables are sub-Gaussian and as noted earlier matrices with i.i.d. Rademacher entries are singular with probability at least $(\frac{1}{2} + o(1))^n$.

As sparse matrices are more abundant in many fields such as statistics, neural network, financial modeling, electrical engineering, wireless communications (we refer the reader to [1, Chapter 7] for further examples, and their relevant references) it is natural to ask if there is an analogue of (1.5)-(1.6) for such matrices. One natural model for sparse random matrices are matrices that are Hadamard products of matrices with i.i.d. entries having a zero mean and unit variance, and matrices with i.i.d. $\text{Ber}(p)$ entries, where $p = o(1)$. In [43] it was shown that (a similar result appeared in [20]) if $p = \Omega(n^{-\alpha})$, for some $\alpha \in (0,1)$ (the notation $a_n = \Omega(b_n)$ implies that $b_n = O(a_n)$), then for such matrices one has that $s_{\min}(A_n) \ge n^{-C_1}$ with large probability, for some large constant $C_1 > 0$. In [20] it was further shown that $s_{\max}(A_n) \le n\sqrt{p}$, with probability approaching one, under a minimal assumption on the moments of the entries. This shows that $\sigma(A_n) = O(n^C)$, for some large constant C, which is much weaker than the prediction (1.6).

In [5], under an optimal moment assumption on the entries, this was improved to show that $\sigma(A_n)$ is indeed $O(n)$ with large probability, whenever $p = \Omega(n^{-\alpha})$ for some $\alpha \in (0,1)$. Results of [5] further show that when the entries of the matrix are products of i.i.d. sub-Gaussian random variables and $\text{Ber}(p)$ variables then $\sigma(A_n) = O(n^{1+o(1)})$ with large probability, as long as $p \ge C\frac{\log n}{n}$, for some large C. This also matches with von Neumann's prediction regarding the condition number of a random matrix except for the factor $n^{o(1)}$. As noted earlier in Corollary 1.3 when p is near $\frac{\log n}{n}$ one starts to see the existence of zero rows and columns, which means that matrix is singular with positive probability, and therefore von Neumann's prediction can no longer hold beyond $\frac{\log n}{n}$ barrier.

In this paper our goal is to show that $\frac{\log n}{n}$ is the sharp threshold where a general class of random matrices with 0/1-valued entries undergoes a transition in their invertibility properties. Moreover, for such matrices we show that the existence of a zero row or a zero column is the main reason for the non-invertibility.

A related research direction was pursued by Costello and Vu in [14] where they analyzed the rank of Adj_n, the adjacency matrix of an Erdős-Rényi graph. Later in [15] they considered the adjacency matrix of an Erdős-Rényi graph with general edge weights. In [14] it was shown that if $np > c \log n$, where $p \in (0,1)$ is the edge connectivity probability and $c > \frac{1}{2}$ is some absolute constant, then the co-rank of Adj_n equals the number of isolated vertices in the graph with probability at least $1 - O((\log \log n)^{-1/4})$. This, in particular establishes an analogue of Corollary 1.3(a)-(b) for such matrices. Since [14] studies only the rank of such matrices, unlike Theorem 1.7 and Corollary 1.16, it does not provide any estimate on the lower bound on $s_{\min}$ and the upper bound on the modified condition number. Moreover, it does not provide an analogue of Corollary 1.3(c).

Before describing the models for the sparse random graphs that we work with in this paper, let us mention the following class of random matrices with 0/1-valued entries that are closely related. Recently, there have been interests to study properties of the adjacency matrices of directed and undirected d-regular random graphs. In the context of the invertibility, it had been conjectured that the adjacency matrices of random d-regular ($d \ge 3$) directed and undirected graphs on n vertices are non-singular with probability approaching one, as $n \to \infty$, see [17, 48, 49]. After a series of partial results [3, 11, 12, 27, 28, 29, 25] the conjecture has been recently proved in [18, 19, 32, 33] for both the configuration model and the permutation model.

The adjacency matrix of a random d-regular graph and that of an Erdős-Rényi graph are similar in nature in many aspects. In the context of the invertibility property, the latter ceases to be non-singular when the average degree drops below $\log n$, and whereas the former remains invertible even when the degree is bounded. As highlighted in Corollary 1.3(c) (see also Remark 1.9) the non-invertibility of the latter is purely due to the existence of a zero row or a column. Since, the former always have d non-zero entries per row and column one does not see the transition in its invertibility property.

Let us now describe the models of the random graphs. We begin with the well known notion of undirected Erdős-Rényi graph.

Definition 1.4 (Undirected Erdős-Rényi graphs)**.** The undirected Erdős-Rényi graph $\mathsf{G}(n,p)$ is a graph with vertex set $[n]$ such that for every pair of vertices i and j the edge between them is present with probability p, independently of everything else. Thus denoting $\mathrm{Adj}(\mathsf{G})$ to be the adjacency matrix of a graph G we see that

$$\mathrm{Adj}(\mathsf{G}(n,p))(i,j) = \begin{cases} \delta_{i,j} & \text{for } i < j \\ \delta_{j,i} & \text{for } i > j \\ 0 & \text{otherwise} \end{cases},$$

where $\{\delta_{i,j}\}_{i<j}$ are i.i.d. $\mathrm{Ber}(p)$ random variables taking one with probability p and zero with probability $(1-p)$.

Next we describe the model for the directed Erdős-Rényi graph.

Definition 1.5 (Directed Erdős-Rényi graphs)**.** We define the directed Erdős-Rényi graph with vertex set $[n]$ as follows: for each pair of vertices i and j the edge between them is drawn with probability $2p$, independently of everything else, and then the direction of the edge is chosen uniformly at random. Such graphs will be denoted by $\overrightarrow{\mathsf{G}}(n,p)$. We therefore note that

$$\mathrm{Adj}(\overrightarrow{\mathsf{G}}(n,p))(i,j) = \begin{cases} \tilde{\delta}_{i,j} \cdot \theta_{i,j} & \text{for } i < j \\ \tilde{\delta}_{j,i} \cdot (1-\theta_{j,i}) & \text{for } i > j \\ 0 & \text{otherwise} \end{cases},$$

where $\{\tilde{\delta}_{i,j}\}_{i<j}$ are i.i.d. $\mathrm{Ber}(2p)$ and $\{\theta_{i,j}\}_{i<j}$ are i.i.d. $\mathrm{Ber}(1/2)$ random variables.

It is easy to note that $\text{Adj}(\overrightarrow{\mathsf{G}}(n,p))$ has the following representation which will be useful later:

$$\text{Adj}(\overrightarrow{\mathsf{G}}(n,p))(i,j) = \begin{cases} \tilde{\delta}_{i,j} \cdot \theta_{i,j} & \text{for } i < j \\ \tilde{\delta}_{i,j} \cdot (1-\theta_{j,i}) & \text{for } i > j \\ 0 & \text{otherwise} \end{cases} \tag{1.11}$$

where $\{\theta_{i,j}\}_{i<j}$ are as above and $\{\tilde{\delta}_{i,j}\}_{i,j=1}^{n}$ are i.i.d. $\text{Ber}(2p)$ random variables. This representation yields additional independence which is exploited in our proofs.

Below we define a random bipartite graph.

Definition 1.6 (Random bipartite graphs)**.** Fix $m, n \in \mathbb{N}$ and let $\mathsf{BG}(m,n,p)$ be a bipartite graph on $[m+n]$ vertices such that for every $i \in [m]$ and $j \in [m+n]\backslash[m]$ the edge between them is present with p, independently of everything else. Therefore,

$$\text{Adj}(\mathsf{BG}(m,n,p))(i,j) = \begin{cases} \delta_{i,j} & \text{for } i \in [m], j \in [n+m]\backslash[m] \\ \delta_{j,i} & \text{for } i \in [n+m]\backslash[m], j \in [m] \\ 0 & \text{otherwise} \end{cases},$$

where $\{\delta_{i,j}\}$ are i.i.d. $\text{Ber}(p)$. When $m = n$, for brevity we write $\mathsf{BG}(n,p)$.

Now we are ready to describe the main result of this paper. Let us recall the definition of Ω_0 from (1.1). Definitions 1.4, 1.5, and 1.6 give rise to three classes of random matrices: fully i.i.d. Bernoulli matrices, symmetric Bernoulli with a zero diagonal, and the third class which does not belong to the classical random matrix theory ensembles. The next theorem states that on the event that the graph has no isolated vertices, the same lower bound for the smallest singular value holds for all three classes.

Theorem 1.7. *Let $A_n = \text{Adj}(\mathsf{G}(n,p))$, $\text{Adj}(\overrightarrow{\mathsf{G}}(n,p))$, or $\text{Adj}(\mathsf{BG}(n,p))$.*

(i) *If $np \geq \log(1/p)$ then there exist absolute constants $0 < c_{1.7}, \widetilde{C}_{1.7}, C_{1.7} < \infty$ such that for any $\varepsilon > 0$, we have*

$$\mathbb{P}\left(\left\{s_{\min}(A_n) \leq c_{1.7}\varepsilon \exp\left(-\widetilde{C}_{1.7}\frac{\log(1/p)}{\log(np)}\right)\sqrt{\frac{p}{n}}\right\} \cap \Omega_0^c\right) \leq \varepsilon^{1/5} + \frac{C_{1.7}}{\sqrt[4]{np}} \tag{1.12}$$

(ii) *If $np \leq \log(1/p)$ then there exists an absolute constant $\bar{C}_{1.7}$ such that*

$$\mathbb{P}(\Omega_0) \geq 1 - \frac{\bar{C}_{1.7}}{\log n}.$$

Remark 1.8. Note that

$$\text{Adj}(\mathsf{BG}(n,p)) := \begin{bmatrix} \mathbf{0}_n & \text{Adj}_{12}(\mathsf{BG}(n,p)) \\ \text{Adj}_{12}^*(\mathsf{BG}(n,p)) & \mathbf{0}_n \end{bmatrix},$$

where $\mathbf{0}_n$ is the $n \times n$ matrix of all zeros and $\text{Adj}_{12}(\mathsf{BG}(n,p))$ is a matrix with i.i.d. $\text{Ber}(p)$ entries. Therefore the set of singular values of $\text{Adj}(\mathsf{BG}(n,p))$ are same with that of $\text{Adj}_{12}(\mathsf{BG}(n,p))$ and each of the singular values of the former has multiplicity two. To simplify the presentation, we will use the $n \times n$ matrix $\text{Adj}_{12}(\mathsf{BG}(n,p))$ as the adjacency matrix of a bipartite graph instead of the $(2n) \times (2n)$ matrix $\text{Adj}(\mathsf{BG}(n,p))$. This is the random matrix with i.i.d. $\text{Ber}(p)$ entries considered above.

Remark 1.9. Theorem 1.7 implies that Corollary 1.3 holds for all three classes of adjacency matrices of random graphs. It means that the phase transition from invertibility to singularity occurs at the same value of p in all three cases, and the main reason for singularity is also the same.

Remark 1.10. A straightforward modification of the proof of Theorem 1.7 shows that the same extends for a symmetric matrix with i.i.d. Bernoulli entries on and above the diagonal. We leave the details to the reader.

Remark 1.11. For $m \sim n$, one can extend the proof of Theorem 1.7 to derive a quantitative lower bound on $s_{\min}(\text{Adj}(\mathsf{BG}(m, n, p)))$. We do not pursue this extension here.

Remark 1.12. A result similar to Theorem 1.7 holds also for general sparse random matrices. We will not discuss it here to keep the paper to a reasonable length.

Building on Theorem 1.7 we now proceed to find an upper bound on the condition number. We point out to the reader that as the entries of A_n have non-zero mean $s_{\max}$ is of larger order of magnitude than the rest of the singular values. So, we cannot expect an analogue of (1.6) to hold for the condition number of A_n where A_n is as in Theorem 1.7. Therefore, we define the following notion of modified condition number.

Definition 1.13. For any matrix A_n we define its modified condition number as follows:

$$\widetilde{\sigma}(A_n) := \frac{s_2(A_n)}{s_{\min}(A_n)}.$$

To obtain an upper bound on $\widetilde{\sigma}(A_n)$ we need the same for $s_2(A_n)$ which follows from the theorem below.

Theorem 1.14. *Let A_n be as in Theorem 1.7. Fix $c_0 > 0, C_0 \geq 1$ and let $p \geq c_0 \frac{\log n}{n}$. Then there exists a constant $C_{1.14}$, depending only on c_0 and C_0 such that*

$$\mathbb{P}\left(\|A_n - \mathbb{E}A_n\| \geq C_{1.14}\sqrt{np}\right) \leq \exp(-C_0 \log n),$$

for all large n.

Remark 1.15. If $A_n = \text{Adj}(\mathsf{G}(n, p_n))$ or $\text{Adj}(\overrightarrow{\mathsf{G}}(n, p_n))$ we note that $p\boldsymbol{J}_n - \mathbb{E}A_n = pI_n$, where $\boldsymbol{J}_n$ is the matrix of all ones and I_n is the identity matrix. Therefore, Theorem 1.14 immediately implies that $\|A_n - p\boldsymbol{J}_n\| = O(\sqrt{np})$ with large probability for such matrices. Since

$$s_2(A_n) = \inf_{v \in \mathbb{R}^n} \sup_{x \in S^{n-1},\, x \perp v} \|A_n x\|_2 \leq \sup_{x \in S^{n-1}} \|(A_n - p\boldsymbol{J}_n)x\|_2,$$

it further yields that the same bound continues to hold for the second largest singular value of the adjacency matrices of directed and undirected Erdős-Rényi graphs. As $\text{Adj}_{12}(\mathsf{BG}(n, p))$ is a matrix with i.i.d. Bernoulli entries we have that $\mathbb{E}\text{Adj}_{12}(\mathsf{BG}(n, p)) = p\boldsymbol{J}_n$. Therefore, recalling Remark 1.8 we deduce from Theorem 1.14 that $s_2(A_n) = O(\sqrt{np})$, with large probability, when A_n is the adjacency matrix of a random bipartite graph.

Remark 1.15 combined with Theorem 1.7 yields the following corollary.

Corollary 1.16. *Let $A_n = \text{Adj}(\mathsf{G}(n, p))$, $\text{Adj}(\overrightarrow{\mathsf{G}}(n, p))$, or $\text{Adj}(\mathsf{BG}(n, p))$. If $np \geq \log(1/p)$ then there exist absolute constants $0 < C_{1.16}, \widetilde{C}_{1.16}, \bar{C}_{1.16} < \infty$ such that for any $\varepsilon > 0$, we have*

$$\mathbb{P}\left(\left\{\widetilde{\sigma}(A_n) \geq C_{1.16}\varepsilon^{-1} n^{1 + \frac{\widetilde{C}_{1.16}}{\log\log n}}\right\} \cap \Omega_0^c\right) \leq \varepsilon^{1/5} + \frac{\bar{C}_{1.16}}{\sqrt[4]{np}}. \tag{1.13}$$

Thus, Corollary 1.16 shows that
up to a set of a small probability, we have a dichotomy: either the matrix A_n contains a zero row or zero column, and so $\widetilde{\sigma}(A_n) = \infty$, or $\widetilde{\sigma}(A_n)$ is roughly of the same order as for the dense random

matrix.
This establishes an analogue of von Neumann's conjecture for the condition number for the entire range of p.

While the results of this paper are formulated for adjacency matrices of graphs, similar statements can be proved for matrices with i.i.d. random entries and for symmetric random matrices whose entries are products of Bernoulli variables and i.i.d. sub-Gaussian variables, i.e., in the setup similar to [5, 50]. We would not discuss this setup here as it would add an additional level of technical arguments to the proof.

The rest of the paper is organized as follows: In Section 2 we provide an outline of the proofs of Theorems 1.7 and 1.14. In Section 3 we show that A_n is well invertible over the set of vectors that are close to sparse vectors. We split the set of such vectors into three subsets: vectors that are close very sparse vectors, close to moderately sparse vectors, and those that have a large spread component. Section 4 shows that the matrix in context is well invertible over the set of vectors that are not close to sparse vectors. In Section 5 we first prove Theorem 1.7(ii) which essentially follows from Markov's inequality. Then combining the results of Sections 3-4 and using Theorem 1.14 we prove Theorem 1.7(i). The proof of Theorem 1.14 can be found in Section 6. Appendix A contains the proofs of some structural properties of the adjacency matrices of the sparse random graphs that are used to treat very sparse vectors. In Appendix B we prove invertibility over vectors that are close to sparse vectors having a large spread component.

2. Proof outline

In this section we provide outlines of the proofs of Theorems 1.7 and 1.14. Broadly, the proof of Theorem 1.14 consists of two parts. One of them is to show that $\|A_n - \mathbb{E}A_n\|$ concentrates near its mean. This is a consequence of Talagrand's concentration inequality for convex Lipschitz functions. The second step is to find a bound on $\mathbb{E}\|A_n - \mathbb{E}A_n\|$. This can be derived using [2]. The proof of Theorem 1.7(ii) follows from standard concentration bounds.

The majority of this paper is devoted to the proof Theorem 1.7(i), i.e. to find a lower bound on the smallest singular value. As we are interested in finding a lower bound on $s_{\min}$ for sparse matrices, we will assume that $p \le c$ for some absolute constant $c \in (0,1)$ whenever needed during the course of the proof.

We begin by noting that

$$s_{\min}(A_n) = \inf_{x \in S^{n-1}} \|A_n x\|_2.$$

To obtain a lower bound on the infimum over the whole sphere we split the sphere into the set of vectors that are close to sparse vectors and its complement. Showing invertibility over these two subsets of the sphere requires two different approaches.

First let us consider the set of vectors that are close to sparse vectors. This set of vectors has a low metric entropy. So, the general scheme would be to show that for any unit vector x that is close to some sparse vector, $\|A_n x\|_2$ cannot be too small with large probability. Then the argument will be completed by taking a union over an appropriate net of the set of such vectors that has a small cardinality.

To obtain an effective probability bound on the event that $\|A_n x\|_2$ is small when x is close to a sparse vector we further need to split the set of such vectors into three subsets: vectors that are close to very sparse vectors, vectors that are close to moderately sparse vectors, and vectors that are close to sparse vectors having a sufficiently large spread component, or equivalently a large non-dominated tail (see Sections 3.1-3.3 for precise formulations).

Unlike the dense set-up, the treatment of very sparse vectors turns out be significantly different for sparse random matrices. It stems from the fact that for such vectors, the small ball probability estimate is too weak to be combined with the union bound over a net. A different method introduced in [5] and subsequently used in [50] relies on showing that for any very sparse vector x, one can find a large sub-matrix of A_n such that it has one non-zero entry per row. It effectively means that there is no cancellation in $(A_n x)_i$ for a large collection of rows $i \in [n]$. This together with the fact that the set of coordinates of x indexed by the columns of the sub-matrix chosen supports a significant proportion of the norm completes the argument. However, as seen in [5], this argument works only when $np \geq C \log n$, for some large constant C. When, $np \leq C \log n$ light columns (i.e. the columns for which the number of non-zero entries is much smaller than np, see also Definition 3.6) start to appear, with large probability. Hence, the above sub-matrix may not exist.

To overcome this obstacle one requires new ideas. Under the current set-up, we show that given any unit vector x, on the event that there is no zero row or column in A_n, the vector $A_n x$ and the coordinates of x that are not included in the set of light columns cannot have a small norm at the same time (see Lemma 3.12). This essentially allows us to look for sub-matrices of A_n having one non-zero entry per row, whose columns do not intersect with the set of light columns. In the absence of the light columns one can use Chernoff bound to obtain such a sub-matrix. This route was taken in [5, 50]. However, as explained above, to carry out the same procedure here we need to condition on events involving light columns of A_n. So the joint independence of the entries is lost and hence Chernoff bound becomes unusable.

To tackle this issue we derive various structural properties of A_n regarding light and normal (i.e. not light) columns. Using this we then show that there indeed exists a large sub-matrix of A_n with desired properties, with large probability. We refer the reader to Lemmas 3.7 and 3.10 for a precise formulation of this step.

Next, we provide an outline of the proof to establish the invertibility over the second and the third sets of sparse vectors. To treat the infimum over such vectors, we first need to obtain small ball probability estimates. This is done by obtaining bounds on the Lévy concentration function which is defined below.

Definition 2.1 (Lévy concentration function)**.** Let Z be a random variable in $\mathbb{R}^n$. For every $\varepsilon > 0$, the Lévy concentration function of Z is defined as

$$\mathcal{L}(Z, \varepsilon) := \sup_{u \in \mathbb{R}^n} \mathbb{P}(\|Z - u\|_2 \leq \varepsilon).$$

The desired bound on the Lévy concentration function for the second set of vectors is a consequence of Paley-Zygmund inequality and a standard tensorization argument. Since the third set of vectors has a higher metric entropy than the second, the small ball probability bound derived for the second set of vectors becomes too weak to take a union bound. So using the fact that any vector belonging to the third set has a large spread component, we obtain a better bound on the Lévy concentration function which is essentially a consequence of the well known Berry-Esséen theorem (see Lemma 3.19). Using this improved bound we then carry out an ε-net argument to show that A_n is also well invertible over the third set of sparse vectors.

Now it remains to provide an outline of the proof of the invertibility over non-sparse vectors. It is well known that such vectors have a large metric entropy, so one cannot use the same argument as above. Instead, using [36] we obtain that it is enough to control $\text{dist}(A_{n,1}, H_{n,1})$, the distance of $A_{n,1}$, the first column of A_n, to $H_{n,1}$, the subspace spanned by the rest of the columns. To control the distance, we derive an expression for it that is more tractable (see Proposition 4.3). From Proposition 4.3, after some preprocessing, we find that it suffices to show that $\langle C_n^{-1} \boldsymbol{x}, \boldsymbol{y} \rangle$ is not too small with large probability, where C_n^{T} is the $(n-1)\times(n-1)$ sub-matrix of A_n obtained by deleting its first row and column, and $\boldsymbol{x}^{\mathsf{T}}$ and $\boldsymbol{y}$ are the first row and column of A_n with the first common entry removed, respectively (if C_n is non-invertible, then there is an alternate and simpler lower bound on the relevant distance).

Since Theorem 1.7 allows $\boldsymbol{x}$ and $\boldsymbol{y}$ to be dependent a bound on $\mathcal{L}(\langle C_n^{-1}\boldsymbol{x},\boldsymbol{y}\rangle,\varepsilon)$ is not readily available. We use a decoupling argument to show that it is enough to find a bound on the Lévy concentration function of the random variable $\langle C_n^{-1}\widehat{\boldsymbol{x}},\widehat{\boldsymbol{y}}\rangle$ for some properly chosen $\widehat{\boldsymbol{x}}$ and $\widehat{\boldsymbol{y}}$, where $\widehat{\boldsymbol{x}}$ and $\widehat{\boldsymbol{y}}$ are now independent. This follows the road-map introduced in [45] for symmetric matrices, although the implementation of it in our case is harder due to the fact that $\boldsymbol{x}$ and $\boldsymbol{y}$ may be different. Having shown this, the desired small ball probability follows once we establish that the random vector $v_\star := C_n^{-1}\widehat{\boldsymbol{x}}$ has a large spread component. Note that $v_\star$ solves the equation $C_n v = \widehat{\boldsymbol{x}}$. We have already established invertibility of C_n over sparse vectors that has a large spread component. Now, we extend that argument to show that any solution of the equation $C_n v = \widehat{\boldsymbol{x}}$ must also have a large spread component. This allows us to deduce the desired properties of $v_\star$. It completes the outline of the proof of Theorem 1.7(i).

3. Invertibility over compressible and dominated vectors

To prove a uniform lower bound on $\|A_n x\|_2$ for x close to sparse vectors when A_n is the adjacency matrix of one of the three models of the random graphs described in Section 1, we will unite them under the following general set-up. It is easy to see that the adjacency matrices of all three models of random graphs satisfy this general assumption.

Assumption 3.1. Let A_n be a $n\times n$ matrix with entries $\{a_{i,j}\}$ such that

(a) The diagonals $\{a_{i,i}\}_{i=1}^n$ and the off-diagonals $\{a_{i,j}\}_{i\neq j}$ are independent of each other.
(b) The random variables $\{a_{i,i}\}_{i=1}^n$ are jointly independent and $a_{i,i}\sim \text{Ber}(p_i)$ with $p_i\le p$ for all $i\in[n]$.
(c) For every $i\neq j\in[n]$, $a_{i,j}\sim\text{Ber}(p)$ and independent of the rest of the entries except possibly $a_{j,i}$.

Remark 3.2. The proofs of the main results of this section extend for matrices with symmetrized Bernoulli entries satisfying the dependency structure of Assumption 3.1. That is, one can consider the matrix A_n with

$$\mathbb{P}(a_{i,j}=\pm 1)=\frac{p}{2},\qquad \mathbb{P}(a_{i,j}=0)=1-p,$$

and $a_{j,i}=-a_{i,j}$. Note that, this extension in particular includes skew-symmetric matrices. Although skew-symmetric matrices of odd dimension are singular, it shows that they are invertible over sparse vectors.

Before proceeding further let us now formally define the notions of vectors that are close to sparse vectors. These definitions are borrowed from [5].

Definition 3.3. Fix $m<n$. The set of m-sparse vectors is given by

$$\text{Sparse}(m) := \{x\in\mathbb{R}^n \mid |\text{supp}(x)|\le m\},$$

where $|S|$ denotes the cardinality of a set S. Furthermore, for any $\delta>0$, the unit vectors which are δ-close to m-sparse vectors in the Euclidean norm, are called (m,δ)-compressible vectors. The set of all such vectors hereafter will be denoted by $\text{Comp}(m,\delta)$. Thus,

$$\text{Comp}(m,\delta) := \{x\in S^{n-1} \mid \exists y\in\text{Sparse}(m) \text{ such that } \|x-y\|_2\le\delta\},$$

where $\|\cdot\|_2$ denotes the Euclidean norm. The vectors in S^{n-1} which are not compressible, are defined to be incompressible, and the set of all incompressible vectors is denoted as $\text{Incomp}(m, \delta)$.

As already seen in [5, 50] for sparse random matrices one can obtain an effective bound over the subset of the incompressible vectors that have a non-dominated tail. This necessitates the following definition of dominated vectors. These are also close to sparse vectors, but in a different sense.

Definition 3.4. For any $x \in S^{n-1}$, let $\pi_x : [n] \to [n]$ be a permutation which arranges the absolute values of the coordinates of x in a non-increasing order. For $1 \le m \le m' \le n$, denote by $x_{[m:m']} \in \mathbb{R}^n$ the vector with coordinates

$$x_{[m:m']}(j) = x_j \cdot \mathbf{1}_{[m:m']}(\pi_x(j)).$$

In other words, we include in $x_{[m:m']}$ the coordinates of x which take places from m to m' in the non-increasing rearrangement.
For $\alpha < 1$ and $m \le n$ define the set of vectors with dominated tail as follows:

$$\text{Dom}(m, \alpha) := \{x \in S^{n-1} \mid \left\|x_{[m+1:n]}\right\|_2 \le \alpha\sqrt{m} \left\|x_{[m+1:n]}\right\|_\infty\}.$$

Note that by definition, $\text{Sparse}(m) \cap S^{n-1} \subset \text{Dom}(m, \alpha)$, since for m-sparse vectors, $x_{[m+1:n]} = 0$.

3.1. **Invertibility over vectors close to very sparse.** As mentioned in Section 2, the key to control the ℓ_2 norm of $A_n x$ when x is close to very sparse vectors is to show that A_n has large sub-matrices containing a single non-zero entry per row. This will be then followed by an ε-net argument and the union bound. As we will see a direct application of this idea requires that $\|A_n\| = O(\sqrt{np})$ which does not hold with high probability, because the entries of A_n have a non-zero mean. To overcome this obstacle we use the *folding trick* introduced in [5].

Definition 3.5 (Folded matrices and vectors)**.** Denote $\mathfrak{n} := \lfloor n/2 \rfloor$. For any $y \in \mathbb{R}^n$ we define

$$\text{fold}(y) := y_1 - y_2,$$

where y_i, $i = 1, 2$, are the vectors in $\mathbb{R}^{\mathfrak{n}}$ whose entries are the first and the next $\mathfrak{n}$ coordinates of y. Similarly for a $n \times n$ matrix B_n we define

$$\text{fold}(B_n) := B_{n,1} - B_{n,2},$$

where $B_{n,i}$, $i = 1, 2$ are $\mathfrak{n} \times n$ matrices consisting of the first and the next $\mathfrak{n}$ rows of B_n.

It is easy to see that except a few of entries of $\text{fold}(A_n)$, the rest have zero mean which allows us to deduce that $\|\text{fold}(A_n)\| = O(\sqrt{np})$ with large probability. Moreover, using the triangle inequality we see that $\|\text{fold}(A_n)x\|_2 \le 2\|A_n x\|_2$. So, we can work with $\text{fold}(A_n)$ instead of A_n.

To obtain the small ball probability estimate on $\|\text{fold}(A_n)x\|_2$, where x is very close to a sparse vector we need to derive some structural properties of A_n.

To this end, we introduce the following notion of light and normal columns and rows.

Definition 3.6 (Light and normal columns and rows)**.** For a matrix B_n and $i \in [n]$, let us write $\text{row}_i(B_n)$ and $\text{col}_i(B_n)$ to denote the i-th row and column of B_n respectively. Let $\delta_0 \in (0, 1/10)$ be a fixed constant. We call $\text{col}_j(B_n)$, $j \in [n]$, light if $|\text{supp}(\text{col}_j((B_n))| \le \delta_0 np$. A column which is not light will be called normal. Similar definitions are adopted for the rows.

Next denote

$$\mathcal{L}(B_n) := \{j \in [n] : \text{col}_j(B_n) \text{ is light}\}.$$

We are now ready to state the following result on the typical structural properties of A_n.

Lemma 3.7 (Structural properties of A_n)**.** *Let A_n satisfies Assumption 3.1 and*

$$np \geq \log(1/\bar{C}p), \tag{3.1}$$

for some $\bar{C} \geq 1$. Denote $\Omega_{3.7}$ be the event such that the followings hold:

(1) (No heavy rows and columns) For any $j \in [n]$,

$$|\operatorname{supp}(\operatorname{row}_j(A_n))|, |\operatorname{supp}(\operatorname{col}_j(A_n))| \leq C_{3.7} np,$$

where $C_{3.7}$ is a large absolute constant.

(2) (Light columns have disjoint supports) For any $(i,j) \in \binom{[n]}{2}$ such that $\operatorname{col}_i(A_n), \operatorname{col}_j(A_n)$ are light, $\operatorname{supp}(\operatorname{col}_i(A_n)) \cap \operatorname{supp}(\operatorname{col}_j(A_n)) = \varnothing$.

(3) (The number of light columns connected to any column is bounded) There is an absolute constant r_0 such that for any $j \in [n]$, the number of light columns $\operatorname{col}_i(A_n)$, $i \in [n]$, with $\operatorname{supp}(\operatorname{col}_i(A_n)) \cap \operatorname{supp}(\operatorname{col}_j(A_n)) \neq \varnothing$ does not exceed r_0.

(4) (The support of a normal column has a small intersection with the light ones) For any $j \in [n]$ such that $\operatorname{col}_j(A_n)$ is normal,

$$\left| \operatorname{supp}(\operatorname{col}_j(\operatorname{fold}(A_n))) \cap \left(\bigcup_{i \in \mathcal{L}(A_n)} \operatorname{supp}(\operatorname{col}_i(\operatorname{fold}(A_n))) \right) \right| \leq \frac{\delta_0}{16} np.$$

(5) (Extension property) For any $I \subset [n]$ with $2 \leq |I| \leq c_{3.7} p^{-1}$

$$\left| \bigcup_{j \in I} (\operatorname{supp}(\operatorname{col}_j(\operatorname{fold}(A_n)))) \right| \geq \sum_{j \in I} |\operatorname{supp}(\operatorname{col}_j(\operatorname{fold}(A_n)))| - \frac{\delta_0}{16} np|I|,$$

where $c_{3.7}$ is a constant depending only on δ_0.

(6) (supports of columns of the matrix and its folded version are close in size) For every $j \in [n]$,

$$\big| |\operatorname{supp}(\operatorname{col}_j(A_n))| - |\operatorname{supp}(\operatorname{col}_j(\operatorname{fold}(A_n)))| \big| \leq \frac{\delta_0}{8} np.$$

Then there exists n_0, depending only $\bar{C}$ and δ_0, such that for any $n \geq n_0$ the event $\Omega_{3.7}$ occurs with probability at least $1 - n^{-\bar{c}_{3.7}}$ for some $\bar{c}_{3.7} > 0$ depending only on δ_0.

The proof of Lemma 3.7 relies on standard tools such as Chernoff bound, and Markov inequality. Its proof is deferred to Appendix A.

Remark 3.8. As we will see in Section 4 (also mentioned in Section 2), to establish the invertibility over incompressible and non-dominated vectors for the adjacency matrices of undirected and directed Erdős-Rényi graphs, one needs to find a uniform lower bound on $\|\operatorname{fold}(A_n)x - y_0\|_2$ over compressible and dominated vectors x and some fixed $y_0 \in \mathbb{R}^n$ with $|\operatorname{supp}(y_0)| \leq C_\star np$ for some $C_\star > 0$. While showing invertibility over vectors that are close to very sparse vectors, we tackle this additional difficulty by deleting the rows from A_n that are in $\operatorname{supp}(y_0)$. This requires proving an analog of Lemma 3.7 for rectangular sub-matrix $\bar{A}_n$ of dimension $\bar{n} \times n$, where $n - C_\star np \leq \bar{n} \leq n$. This means that to apply Lemma 3.7 for the original matrix A_n we need to prove it under the assumption (3.1) rather than the assumption $np \geq \log(1/p)$. To keep the presentation of this paper simpler we prove Lemma 3.7 only for square matrices. Upon investigating the proof it becomes clear that the extension to rectangular, almost square, matrices requires only minor changes.

Next we define the following notion of a good event needed to establish the small ball probability estimates on $\operatorname{fold}(A_n)x - \operatorname{fold}(y_0)$ for x close to very sparse vectors and some fixed vector $y_0 \in \mathbb{R}^n$.

Definition 3.9 (Good event)**.** Let A_n satisfy Assumption 3.1. Fix $\kappa \in \mathbb{N}$, $J, J' \in [n]$ disjoint sets. Denote

$$\bar{J} := \bar{J}(J) := \{j \in [\mathfrak{n}] : j \in J \text{ or } j + \mathfrak{n} \in J\},$$

and similarly $\bar{J}'$. For any $c > 0$, define $\mathcal{A}_c^{J,J'}$ to be the event that there exists $I \subset [\mathfrak{n}] \backslash (\bar{J} \cup \bar{J}')$ with $|I| \ge c\kappa np$ such for every $i \in I$ there further exists $j_i \in J$ so that

$$|\mathfrak{a}_{i,j_i}| = 1, \quad \mathfrak{a}_{i,j} = 0 \text{ for all } j \in J \cup J' \backslash \{j_i\},$$

where $\{\mathfrak{a}_{i,j}\}$ are the entries of $\text{fold}(A_n)$, and

$$\text{supp}(\text{row}_i(\text{fold}(A_n))) \cap \mathcal{L}(A_n) = \varnothing, \quad \text{for all } i \in I. \tag{3.2}$$

Now we are ready to state the structural lemma that shows that the good event $\mathcal{A}_c^{J,J'}$ holds with high probability for appropriate sizes of J and J'.

Lemma 3.10. *Let A_n satisfy Assumption 3.1 and $np \ge \log(1/\bar{C}p)$ for some $\bar{C} \ge 1$. Then, there exist an absolute constant $\bar{c}_{3.10}$, and constants $c_{3.10}, c^\star_{3.10}$, depending only on δ_0, such that*

$$\mathbb{P}\left(\bigcup_{\kappa \le c^\star_{3.10}(p\sqrt{pn})^{-1} \vee 1} \; \bigcup_{\substack{J \in \binom{[n]}{\kappa} \\ J \cap \mathcal{L}(A_n) = \varnothing}} \; \bigcup_{J' \in \binom{[n]}{\mathfrak{m}},\, J \cap J' = \varnothing} \left(\mathcal{A}_{c_{3.10}}^{J,J'}\right)^c \cap \Omega_{3.7} \right) \le n^{-\bar{c}_{3.10}}, \tag{3.3}$$

for all large n, where for $\kappa \in \mathbb{N}$ we write

$$\mathfrak{m} = \mathfrak{m}(\kappa) := \kappa\sqrt{pn} \wedge \frac{c^\star_{3.10}}{p}.$$

Remark 3.11. We point out to the reader that [5, Lemma 3.2] derives a result similar to Lemma 3.10. The key difference is that the former assumes $np \ge C \log n$, for some large constant C, which allows to use Chernoff bound to conclude that given any set of columns $J \subset [n]$ of appropriate size, there is a large number of rows for which there exists exactly one non-zero entry per row in the columns indexed by J. When $np \le C \log n$ this simply does not hold for all $J \subset [n]$ as there are light columns. Moreover, for such choices of p the Chernoff bound is too weak to yield any non-trivial bound on the number of rows with the desired property. Therefore we need to use several structural properties of our matrix A_n, derived in Lemma 3.7, to obtain a useful lower bound on the number of such rows.

Proof of Lemma 3.10. Fixing $\kappa \le c^\star_{3.10}(p\sqrt{pn})^{-1}$, for some constant $c^\star_{3.10}$ to be determined during the course of the proof, we let $J \in \binom{[n]}{\kappa}$. Let $I^1(J)$ be the set of all rows of $\text{fold}(A_n)$ containing exactly one non-zero entry in the columns corresponding to J. More precisely,

$$I^1(J) := \Big\{ i \in [\mathfrak{n}] : \mathfrak{a}_{i,j_i} \ne 0 \text{ for some } j_i \in J, \text{ and } \mathfrak{a}_{i,j} = 0 \text{ for all } j \in J \backslash \{j_i\} \Big\}.$$

Similarly for a set $J' \in \binom{[n]}{m}$ we define

$$I^0(J') := \Big\{ i \in [\mathfrak{n}] \backslash (\bar{J} \cup \bar{J}') : \mathfrak{a}_{i,j} = 0 \text{ for all } j \in J' \Big\}.$$

Note that we have deleted the rows in $\bar{J} \cup \bar{J}'$ while defining $I^0(J')$. This is due to the fact that matrices satisfying Assumption 3.1 allow some dependencies among its entries. Later, in the proof we will require $I^1(J)$ and $I^0(J')$ to be independent for disjoint J and J'.

To estimate $|I^1(J)|$ we let $\mathscr{E} := (\cup_{j\in J} \operatorname{supp}(\operatorname{col}_j(\operatorname{fold}(A_n))))$ and define a function $f : \mathscr{E} \to \mathbb{N}$ by

$$f(i) := \sum_{j\in J} \mathbb{I}\{i \in \operatorname{supp}(\operatorname{col}_j(\operatorname{fold}(A_n)))\}, \quad i \in \mathscr{E}.$$

We note that $I^1(J) = \{i \in \mathscr{E} : \ f(i) = 1\}$. Hence,

$$|I^1(J)| \ge 2|\mathscr{E}| - \sum_{i\in\mathscr{E}} f(i) = \sum_{i\in\mathscr{E}} f(i) - 2\left(\sum_{i\in\mathscr{E}} f(i) - |\mathscr{E}|\right). \tag{3.4}$$

If $J \cap \mathscr{L}(A_n) = \varnothing$, then by property (6) of the event $\Omega_{3.7}$ we have

$$\sum_{i\in\mathscr{E}} f(i) = \sum_{i\in J} |\operatorname{supp}(\operatorname{col}_j(\operatorname{fold}(A_n)))| \ge \sum_{i\in J} |\operatorname{supp}(\operatorname{col}_j(A_n))| - \frac{\delta_0}{8}np|J| \ge \frac{7\delta_0}{8}np|J|.$$

Thus, upon using property (5) of the event $\Omega_{3.7}$, it follows that

$$\sum_{i\in\mathscr{E}} f(i) - |\mathscr{E}| = \sum_{i\in J} |\operatorname{supp}(\operatorname{col}_j(A_n))| - |\mathscr{E}| \le \frac{\delta_0}{16}np|J|.$$

Therefore, from (3.4) we deduce that

$$|I^1(J)| \ge \frac{\delta_0}{2}np|J| \tag{3.5}$$

on the event $\Omega_{3.7}$.

Using the above lower bound on the cardinality of $I^1(J)$ we now show that it has a large intersection with $I^0(J')$. Therefore we can set the desired collection of rows to be the intersection of $I^1(J)$ and $I^0(J')$. However, the caveat with this approach is that the collection of rows just described does not satisfy the property (3.2). To take care of this obstacle, we define

$$\bar{I}^1(J) := I^1(J)\backslash\mathscr{T}(A_n), \quad \text{where} \quad \mathscr{T}(A_n) := \bigcup_{j\in\mathcal{L}(A_n)} \operatorname{supp}(\operatorname{col}_j(\operatorname{fold}(A_n)))$$

From the definition of $\mathscr{T}(A_n)$ it is evident that any subset $I \subset \bar{I}^1(J)$ now satisfies the property (3.2). We further note that

$$|I^1(J) \cap \mathscr{T}(A_n)| \le \sum_{j\in J} |\operatorname{supp}(\operatorname{col}_j(\operatorname{fold}(A_n))) \cap \mathscr{T}(A_n)| \le \frac{\delta_0}{16}np|J|,$$

where in the last step we have used the property (4) of $\Omega_{3.7}$. Thus we proved that on the event $\Omega_{3.7}$,

$$|\bar{I}^1(J)| \ge \frac{7\delta_0}{16}np|J| \tag{3.6}$$

for any $J \subset [n]$ satisfying $J \cap \mathscr{L}(A_n) = \varnothing$.

It remains to show that $I^0(J') \cap \bar{I}^1(J)$ has a large cardinality with high probability. To prove it, we recall that $I^0(J') \subset [\mathfrak{n}]\backslash(\bar{J} \cap \bar{J}')$. Thus, using Assumption 3.1 we find that for any $J' \subset \binom{[n]}{\mathfrak{m}}$ and any $i \in [\mathfrak{n}]\backslash(\bar{J} \cup \bar{J}')$

$$\mathbb{P}(i \in I^0(J')) \ge (1-2p)^{|J|} \ge 1 - 2p\mathfrak{m}.$$

Hence, for a given $I \subset [\mathfrak{n}]$, $\mathbb{E}|I\backslash I^0(J')| \le 2p\mathfrak{m} \cdot |I| \le |I|/4$ by the assumptions on κ and $\mathfrak{m}$. So, by Chernoff's inequality

$$\mathbb{P}\left(|I\backslash I^0(J')| \ge \frac{1}{2}|I|\right) \le \exp\left(-\frac{|I|}{16}\log\left(\frac{1}{8p\mathfrak{m}}\right)\right).$$

Therefore, for any $I \subset [\mathfrak{n}]$ such that $|I| \geq \frac{\delta_0}{4}\kappa np$, we deduce that

$$\mathbb{P}\Big(\exists J' \in \binom{[n]}{m} \text{ such that } |I^0(J') \cap I| \leq \frac{\delta_0}{8}\kappa np\Big)$$

$$\leq \sum_{J' \in \binom{[n]}{\mathfrak{m}}} \mathbb{P}(|I \backslash I^0(J')| \geq \frac{1}{2}|I|)$$

$$\leq \binom{n}{\mathfrak{m}} \cdot \exp\left(-\frac{|I|}{16}\log\left(\frac{1}{4p\mathfrak{m}}\right)\right)$$

$$\leq \exp\left(\mathfrak{m}\cdot\log\left(\frac{en}{\mathfrak{m}}\right) - \frac{\delta_0}{64}\kappa np \cdot \log\left(\frac{1}{8p\mathfrak{m}}\right)\right) = \exp(-\kappa np \cdot U),$$

where

$$U := \frac{\delta_0}{64}\log\left(\frac{1}{8p\mathfrak{m}}\right) - \frac{\mathfrak{m}}{\kappa np}\log\left(\frac{en}{\mathfrak{m}}\right).$$

We now need to find a lower bound on U for which we split the ranges of p. First let us consider $p \in (0,1)$ such that $c^\star_{3.10}(p\sqrt{np})^{-1} \geq 1$. For such choices of p we will show that for any $\kappa \leq c^\star_{3.10} \cdot (p\sqrt{np})^{-1}$, with $c^\star_{3.10}$ sufficiently small, and $\mathfrak{m} = \kappa\sqrt{np}$ we have $U \geq 3$. To this end, denote

$$\alpha := \frac{1}{8\kappa p\sqrt{pn}} \geq \frac{1}{8c^\star_{3.10}}. \tag{3.7}$$

Note that

$$U = \frac{\delta_0}{64}\log\left(\frac{1}{8\kappa p\sqrt{np}}\right) - \frac{1}{\sqrt{np}}\log\left(\frac{en}{\kappa\sqrt{np}}\right) = \frac{\delta_0}{64}\log\alpha - \frac{1}{\sqrt{np}}\log(8epn\alpha)$$
$$= \frac{\delta_0}{64}\log\alpha - \frac{1}{\sqrt{np}}\log\alpha - \frac{1}{\sqrt{pn}}\big(\log(8e) + \log(np)\big)$$
$$\geq \frac{\delta_0}{128}\log\alpha,$$

for all large n, where the last step follows upon noting that by the assumption on p we have $np \to \infty$ as $n \to \infty$, using the fact that $x^{-1/2}\log x \to 0$ as $x \to \infty$, and the lower bound on α (see (3.7)). Choosing $c^\star_{3.10}$ sufficiently small and using (3.7) again we deduce that $U \geq 3$, for all large n.

Now let us consider $p \in (0,1)$ such that $c^\star_{3.10}(p\sqrt{np})^{-1} < 1$. For such choices of p we have that $\kappa = 1$ and $\mathfrak{m} = c^\star_{3.10}p^{-1}$. Therefore, recalling the definition of U we note that

$$U = \frac{\delta_0}{64}\log\left(\frac{1}{8c^\star_{3.10}}\right) - \frac{c^\star_{3.10}}{np^2}\log(e(c^\star_{3.10})^{-1}np) \geq \frac{\delta_0}{64}\log\left(\frac{1}{8c^\star_{3.10}}\right) - (c^\star_{3.10})^{-1/3}n^{-1/3}\log(e(c^\star_{3.10})^{-1}n)$$
$$\geq \frac{\delta_0}{128}\log\left(\frac{1}{8c^\star_{3.10}}\right) \geq 3,$$

where the first inequality follows from the assumption that $c^\star_{3.10}(p\sqrt{np})^{-1} < 1$.

This proves that, for any $p \in (0,1)$ such that $np \geq \log(1/\bar{C}p)$ and any $I \subset [\mathfrak{n}]$ with $|I| \geq \frac{\delta_0}{4}\kappa np$, we have

$$\mathbb{P}\left(\exists J' \in \binom{[n]}{\mathfrak{m}} \text{ such that } |I^0(J') \cap I| \leq \frac{\delta_0}{8}\kappa np\right) \leq \exp(-3\kappa np).$$

To finish the proof, for a set $J \in \binom{[n]}{\kappa}$ we define

$$p_J := \mathbb{P}\left(\left\{\exists J' \in \binom{[n]}{\mathfrak{m}} \text{ such that } J' \cap J = \varnothing,\ |\bar{I}^1(J) \cap I^0(J')| < \frac{\delta_0}{8}\kappa np\right\} \cap\ \Omega_{3.7}\right).$$

Since J and J' are disjoint and $I^0(J') \subset [\mathfrak{n}]\setminus(\bar{J} \cup \bar{J}')$, it follows from Assumption 3.1 that the random subsets $\bar{I}^1(J)$ and $I^0(J')$ are independent. Using (3.6) we obtain that for $J \subset [n]$ such that $J \cap \mathcal{L}(A_n) = \varnothing$,

$$p_J \le \sum_{I\subset[\mathfrak{n}],\ |I|>\frac{\delta_0}{4}\kappa np} \mathbb{P}(\bar{I}^1(J)=I)\mathbb{P}\Big(\exists J' \in \binom{[n]}{\mathfrak{m}} \text{ such that } |I^0(J')\cap I| \le \frac{\delta_0}{8}\kappa np\Big)$$

$$\le \exp(-3\kappa np) \sum_{I\subset[\mathfrak{n}],\ |I|>\frac{\delta_0}{4}\kappa np} \mathbb{P}(\bar{I}^1(J)=I) \le \exp(-3\kappa np), \tag{3.8}$$

for all large n. The rest of the proof consists of taking union bounds. First, using the union bound over $J \in \binom{[n]}{\kappa}$ satisfying $J \cap \mathcal{L}(A_n) = \varnothing$, setting $c_{3.10} = \delta_0/16$, we get that

$$\mathbb{P}\left(\bigcup_{\substack{J\in\binom{[n]}{\kappa} \\ J\cap\mathcal{L}(A_n)=\varnothing}} \bigcup_{J'\in\binom{[n]}{\mathfrak{m}},\ J\cap J'=\varnothing} (\mathcal{A}^{J,J'}_{c_{3.10}})^c \cap \Omega_{3.7} \right) \le \binom{n}{\kappa}\exp(-3\kappa np) \le \exp(\kappa\log n - 3\kappa np)$$
$$\le \exp(-\kappa np).$$

Finally taking another union bound over $\kappa \le c^\star_{3.10}(p\sqrt{np})^{-1} \vee 1$ we obtain the desired result. □

Note that in (3.3) we could only consider $J \subset [n]$ such that $J \cap \mathcal{L}(A_n) = \varnothing$. As we will see later, when we apply Lemma 3.10 to establish the invertibility over vectors that are close to sparse, we have to know that $\|A_n x_{[n]\setminus\mathcal{L}(A_n)}\|_2$ is large for $x \in S^{n-1}$ close to very sparse vectors. So, one additionally needs to show that $\|A_n x\|_2$ and $\|x_{[n]\setminus\mathcal{L}(A_n)}\|_2$ cannot be small at the same time. The following lemma does this job. Its proof again uses the structural properties of A_n derived in Lemma 3.7.

Before stating the next lemma let us recall that Ω_0^c is the event that the matrix A_n has neither zero columns nor zero rows (see (1.1)).

Lemma 3.12. *Let A_n satisfy Assumption 3.1. Fix a realization of A_n such that the event $\Omega_0^c \cap \Omega_{3.7}$ occurs. Let $x \in S^{n-1}$ be such that $\|A_n x\|_2 < 1/4$. Then*

$$\left\|x_{[n]\setminus\mathcal{L}(A_n)}\right\|_2 \ge \frac{1}{C_{3.12}np},$$

for some absolute constant $C_{3.12}$.

Proof. We may assume that $\left\|x_{\mathcal{L}(A_n)}\right\|_2 \ge 1/2$, since otherwise there is nothing to prove. For any $j \in \mathcal{L}(A_n)$, we choose $i := i_j$ such that $a_{i_j,j} = 1$. Such a choice is possible since we have assumed that Ω_0^c occurs. Using the property (2) of the event $\Omega_{3.7}$ we see that the function $i : \mathcal{L}(A_n) \to [n]$ is an injection.

Set

$$J_0 := \{j \in \mathcal{L}(A_n) :\ |(A_n x)_{i_j}| \ge (1/2)|x_j|\}.$$

If $\|x_{J_0}\|_2 \ge 1/2$, then

$$\|A_n x\|_2 \ge \left(\sum_{j\in J_0} ((A_n x)_{i_j})^2\right)^{1/2} \ge \frac{1}{4},$$

which contradicts our assumption $\|A_n x\|_2 < \frac{1}{4}$. Hence, denoting $J_1 := \mathcal{L}(A_n)\backslash J_0$, we may assume that $\|x_{J_1}\|_2 \ge 1/4$. We then observe that for any $j \in J_1$,

$$\begin{aligned}\frac{1}{2}|x_j| \ge |(A_n x)_{i_j}| &\ge |a_{i_j,j} x_j| - \left|\sum_{k\neq j} a_{i_j,k} x_k\right| \\ &\ge |x_j| - |\operatorname{supp}(\text{row}_{i_j}(A_n))| \cdot \max_{k\in\operatorname{supp}(\text{row}_{i_j}(A_n))\backslash\{j\}} |x_k| \\ &\ge |x_j| - C_{3.7} np \cdot \max_{k\in\operatorname{supp}(\text{row}_{i_j}(A_n))\backslash\{j\}} |x_k|,\end{aligned}$$

where the last inequality follows upon using the property (1) of the event $\Omega_{3.7}$.

This shows that for any $j \in J_1$ there exists $k \in [n]$ such that $a_{i_j,k} = 1$ and

$$|x_k| \ge \frac{1}{2C_{3.7}np}|x_j|.$$

Choose one such k and denote it by $k(j)$. Using the property (2) of the event $\Omega_{3.7}$ again, we deduce that $k(j) \in [n]\backslash\mathcal{L}(A_n)$. Therefore,

$$\frac{1}{16} \le \sum_{j\in J_1} x_j^2 \le (2C_{3.7}np)^2 \sum_{j\in J_1} x_{k(j)}^2 \le (2C_{3.7}np)^2 r_0 \sum_{k\in[n]\backslash\mathcal{L}(A_n)} x_k^2.$$

Here, the last inequality follows since by (3) of the event $\Omega_{3.7}$, we have that for any $k \in [n]\backslash\mathcal{L}(A_n)$,

$$|\{j \in \mathcal{L}(A_n) : \ k(j) = k\}| \le r_0.$$

This finishes the proof of the lemma. □

We see that Lemma 3.12 provides a lower bound on $\|x_{[n]\backslash\mathcal{L}(A_n)}\|_2$ that deteriorates as p increases. We show below that for large p the set of light columns is empty with high probability. Hence, in that regime we can work with x instead of $x_{[n]\backslash\mathcal{L}(A_n)}$. Furthermore, during the course of the proof of Proposition 3.14 we will see that to deduce that $x_{[n]\backslash\mathcal{L}(A_n)}$ itself is close to sparse vectors we need bounds on $|\mathcal{L}(A_n)|$ for all p satisfying $np \ge \log(1/p)$. Both these statements are proved in the following lemma.

Lemma 3.13. *Let A_n satisfies Assumption 3.1. If $np \ge \log(1/\bar{C}p)$ for some $\bar{C} \ge 1$ then*

$$\mathbb{P}(|\mathcal{L}(A_n)| \ge n^{\frac{1}{3}}) \le n^{-\frac{1}{9}},$$

for all large n. Moreover, there exists an absolute constant $C_{3.13}$ such that if $np \ge C_{3.13}\log n$ then

$$\mathbb{P}(\mathcal{L}(A_n) = \varnothing) \ge 1 - 1/n.$$

Proof of Lemma 3.13 follows from standard concentration bounds and is postponed to Appendix A. Equipped with all the relevant ingredients we are now ready to state the main result of this section.

Proposition 3.14 (Invertibility over very sparse vectors)**.** *Let A_n satisfies Assumption 3.1 where p satisfies the inequality*

$$np \ge \log(1/p).$$

Fix $K, C_\star \geq 1$ and let

$$\ell_0 := \left\lceil \frac{\log\left(\frac{c^\star_{3.10}}{p}\right)}{\log \sqrt{pn}} \right\rceil. \tag{3.9}$$

Then there exist constants $0 < c_{3.14}, \widetilde{c}_{3.14} < \infty$, depending only on δ_0, and an absolute constant $\bar{c}_{3.14}$ such that for any $y_0 \in \mathbb{R}^n$ with $|\operatorname{supp}(y_0)| \leq C_\star np$, for some $C_\star > 0$, we have

$$\mathbb{P}\left(\{\exists x \in V_0 \text{ such that } \|A_n x - y_0\|_2 \leq \rho\sqrt{np} \text{ and } \|A_n - \mathbb{E}A_n\| \leq K\sqrt{np}\} \cap \Omega_0^c \cap \Omega_{3.7}\right) \leq n^{-\bar{c}_{3.14}},$$

for all large n, where

$$V_0 := \operatorname{Dom}\left(c^\star_{3.10} p^{-1}, c_{3.14} K^{-1}\right) \cup \operatorname{Comp}(c^\star_{3.10} p^{-1}, \rho), \quad \text{and} \quad \rho := (\widetilde{c}_{3.14}/K)^{2\ell_0+1}. \tag{3.10}$$

Proof of Proposition 3.14. Since we do not have any control on the vector y_0 except the cardinality of its support, to remove the effect of y_0 we will show that the ℓ_2-norm of the vector $A_n x - y_0$ restricted to the complement of $\operatorname{supp}(y_0)$ has a uniform bound over $x \in V_0$. To this end, for ease of writing, we define $\bar{A}_n$ to be the sub-matrix of A_n of dimension $\bar{n} \times n$, where $\bar{n} := n - |\operatorname{supp}(y_0)|$, obtained by deleting the rows in $\operatorname{supp}(y_0)$. We have that $\|\bar{A}_n x\|_2 \leq \|A_n x - y_0\|_2$.

Next we observe that for any $x \in \mathbb{R}^n$ an application of the triangle inequality implies that

$$\|\operatorname{fold}(\bar{A}_n)x\|_2^2 \leq 2\|\bar{A}_n x\|_2^2.$$

Furthermore

$$\begin{aligned} \|\operatorname{fold}(\bar{A}_n)\| &\leq \|\bar{A}_{n,1} - \mathbb{E}\bar{A}_{n,1}\| + \|\bar{A}_{n,2} - \mathbb{E}\bar{A}_{n,2}\| + \|\mathbb{E}\bar{A}_{n,1} - \mathbb{E}\bar{A}_{n,2}\| \\ &\leq 2\|A_n - \mathbb{E}A_n\| + \|\mathbb{E}A_{n,1} - \mathbb{E}A_{n,2}\| \leq 2\|A_n - \mathbb{E}A_n\| + 2\sqrt{np}, \end{aligned} \tag{3.11}$$

where in the last step we have used the fact that

$$\|\mathbb{E}A_{n,1} - \mathbb{E}A_{n,2}\| \leq \sqrt{2n} \cdot p \leq 2\sqrt{np}. \tag{3.12}$$

To establish (3.12) we note that Assumption 3.1 implies that there at most two non-zero entries per row in the matrix $\mathbb{E}A_{n,1} - \mathbb{E}A_{n,2}$ each of which has absolute value less than or equal to p. Therefore, each of the entries of $(\mathbb{E}A_{n,1} - \mathbb{E}A_{n,2})(\mathbb{E}A_{n,1} - \mathbb{E}A_{n,2})^*$ is bounded by $2p^2$ and hence by the Gershgorin circle theorem we deduce (3.12).

Therefore, in light of (3.11), recalling $K \geq 1$, it is enough to find a bound on the probability of the event

$$\mathfrak{V}_{V_0} := \left\{\exists x \in V_0 : \left\|\operatorname{fold}(\bar{A}_n)x\right\|_2 \leq 2(\widetilde{c}_{3.14}/K)^{2\ell_0+1}\sqrt{np}\right\} \cap \Omega_K \cap \Omega_0^c \cap \Omega_{3.7},$$

where

$$\Omega_K := \{\|\operatorname{fold}(A_n)\| \leq 4K\sqrt{np}\}.$$

We will show that

$$\mathbb{P}\Bigg(\left\{\exists x \in \operatorname{Dom}\left(c^\star_{3.10}p^{-1}, c_{3.14}K^{-1}\right) : \left\|\operatorname{fold}(\bar{A}_n)x\right\|_2 \leq (\widetilde{c}_{3.14}/K)^{2\ell_0}\sqrt{np}\right\} \cap \Omega_K \cap \Omega_0^c \cap \Omega_{3.7}\Bigg) \leq n^{-\bar{c}_{3.14}}. \tag{3.13}$$

First let us show that $\mathbb{P}(\mathfrak{V}_{V_0}) \le n^{-\bar{c}_{3.14}}$ assuming (3.13). To this end, denoting $m := c^\star_{3.10} p^{-1}$, we note that for any $x \in \mathrm{Comp}(m,\rho)$

$$\left| \| \operatorname{fold}(\bar{A}_n) x\|_2 - \left\| \operatorname{fold}(\bar{A}_n) \frac{x_{[1:m]}}{\|x_{[1:m]}\|_2} \right\|_2 \right| \le \| \operatorname{fold}(\bar{A}_n)\| \cdot \left(\left\| x_{[1:m]} - \frac{x_{[1:m]}}{\|x_{[1:m]}\|_2} \right\|_2 + \|x_{[m+1:n]}\|_2 \right)$$
$$\le 4K\sqrt{np} \cdot \left(1 - \|x_{[1:m]}\|_2 + \|x_{[m+1:n]}\|_2\right)$$
$$\le 8K\rho\sqrt{np} = 8\widetilde{c}_{3.14} \cdot (\widetilde{c}_{3.14}/K)^{2\ell_0} \sqrt{np}, \tag{3.14}$$

on the event Ω_K. For $x \in V_0$ we have that $x_{[1:m]}/\|x_{[1:m]}\|_2 \in \mathrm{Sparse}(m) \cap S^{n-1} \subset \mathrm{Dom}\left(m, c_{3.14}K^{-1}\right)$ we see from (3.13) that

$$\left\| \operatorname{fold}(\bar{A}_n) \frac{x_{[1:m]}}{\|x_{[1:m]}\|_2} \right\|_2 \ge (\widetilde{c}_{3.14}/K)^{2\ell_0} \sqrt{np}$$

with the desired high probability. Therefore, upon shrinking $\widetilde{c}_{3.14}$ such that $10\widetilde{c}_{3.14} \le 1$, and recalling that $K \ge 1$, we deduce from (3.14) that $\mathbb{P}(\mathfrak{V}_{V_0}) \le n^{-\bar{c}_{3.14}}$. Thus, it now suffices to prove (3.13).

Turning to this task, we split the proof into three parts depending on the sparsity level of the matrix A_n, determined by p. First let us consider the case $\log(1/p) \le np \le C_{3.13} \log n$.

Fix $x \in \mathrm{Dom}\left(m, c_{3.14}K^{-1}\right)$ and define $\hat{x} \in \mathbb{R}^n$ to be normalized vector obtained from x after setting the coordinates belonging to the set $\mathcal{L}(\bar{A}_n)$ to be zero. That is,

$$\hat{x}_i := \frac{x_i}{\left\| x_{[n]\setminus \mathcal{L}(\bar{A}_n)} \right\|_2} \cdot \mathbb{I}(i \in [n]\setminus\mathcal{L}(\bar{A}_n)),\ i \in [n].$$

Let us rearrange the magnitudes of the coordinates of $\hat{x}$ and group them in blocks of lengths $(pn)^{\ell/2}$, where $\ell = 1, \ldots, \ell_0$. More precisely, set

$$\hat{z}_\ell := \hat{x}_{[(pn)^{(\ell-1)/2}+1:(pn)^{\ell/2}]}{}^{1}, \tag{3.15}$$

and

$$\hat{z}_{\ell_0+1} := \hat{x}_{[(pn)^{\ell_0/2}+1:n]}. \tag{3.16}$$

For clarity of presentation, let us assume that $m = (pn)^{\ell_0/2}$, i.e. the integer part in the definition of ℓ_0 is redundant. Further denote $\tilde{x} := \hat{x} \cdot \|x_{[n]\setminus\mathcal{L}(\bar{A}_n)}\|_2$. That is, $\tilde{x}$ is the unnormalized version of $\hat{x}$. Note that $\tilde{x}$ matches with x except $|\mathcal{L}(\bar{A}_n)|$ coordinates. Therefore, for any $x \in \mathrm{Dom}(m, c_{3.14}K^{-1})$ we find that

$$\|\tilde{x}_{[m+1:n]}\|_2 \le \|x_{[m+1:n]}\|_2 \le c_{3.14}K^{-1}\sqrt{m}\|x_{[m+1:n]}\|_\infty$$
$$\le \sqrt{\frac{m}{m/2 - |\mathcal{L}(\bar{A}_n)|}} \cdot c_{3.14}K^{-1}\|x_{[m/2+|\mathcal{L}(\bar{A}_n)|+1:m]}\|_2$$
$$\le 2c_{3.14}K^{-1}\|\tilde{x}_{[m/2+1:m]}\|_2,$$

on the event

$$\Omega_{\mathcal{L}} := \left\{ |\mathcal{L}(\bar{A}_n)| \le n^{1/3} \right\},$$

for all large n, where in the last step we have used the fact that for $p = O(\frac{\log n}{n})$ we have $n^{1/3} = o(m)$. This further implies that

$$\|\hat{z}_{\ell_0+1}\|_2 \le 2c_{3.14}K^{-1}\|\hat{x}_{[m/2+1:m]}\|_2 \le 2c_{3.14}K^{-1}\|\hat{z}_{\ell_0}\|_2, \tag{3.17}$$

[1] when $\ell = 1$ by a slight abuse of notation we take $\hat{z}_1 = x_{[1:\sqrt{np}]}$.

on the event $\Omega_{\mathcal{L}}$, where the last inequality is a consequence of the fact that the condition $np \to \infty$ as $n \to \infty$ implies that the support of $\hat{z}_{\ell_0}$ contains that of $\hat{x}_{[m/2+1:m]}$.

Since $\sum_{\ell=1}^{\ell_0+1} \|\hat{z}_\ell\|_2^2 = 1$, we deduce from (3.17) that

$$\sum_{\ell=1}^{\ell_0} \|\hat{z}_\ell\|_2^2 \geq 1 - 4c_{3.14}^2 K^{-2}.$$

Hence, choosing $c_{3.14}$ sufficiently small we obtain that there exists $\ell \leq \ell_0$ such that $\|\hat{z}_\ell\|_2 \geq (c_{3.14}/K)^\ell$. Let $\ell_\star$ be the largest index having this property, and set $u := \sum_{\ell=1}^{\ell_\star} \hat{z}_\ell$, $v =: \sum_{\ell=\ell_\star+1}^{\ell_0+1} \hat{z}_\ell$. First consider the case when $\ell_\star < \ell_0$. Then by the triangle inequality we have that

$$\|v\|_2 \leq \sum_{m=\ell_\star+1}^{\ell_0+1} \|\hat{z}_m\|_2 \leq 4(c_{3.14}/K)^{(\ell_\star+1)}, \tag{3.18}$$

where we have used the inequality (3.17).

Let $\kappa = (np)^{(\ell_\star-1)/2}$. Note that

$$\kappa \leq (np)^{(\ell_0-1)/2} \leq \frac{1}{c_{3.10}^\star p\sqrt{pn}}.$$

To finish the proof we now apply Lemma 3.10 with this choice of κ. Using the fact that $|\operatorname{supp}(y_0)| \leq C_\star np$ we see that

$$\bar{n}p \geq \log(1/\bar{C}p), \tag{3.19}$$

for some large constant $\bar{C}$, whenever $p \leq c$ for some small constant c. Therefore, we can apply Lemma 3.10 to the rectangular matrix $\bar{A}_n$ to find the desired uniform bound on $\|\operatorname{fold}(\bar{A}_n)x\|_2$. To this end, we split the support of u into $\sqrt{np}$ blocks of equal size κ and define $L_{\ell_\star} := \pi_{\hat{x}}^{-1}([1,(np)^{\ell_\star/2}])$, where $\pi_{\hat{x}}$ is the permutation of absolute values of the coordinates of $\hat{x}$ in an non-increasing order. For $s \in [\sqrt{np}]$, define $J_s := \pi_{\hat{x}}^{-1}([(s-1)\kappa+1, s\kappa])$, and set $J'_s := L_{\ell_\star}\backslash J_s$. Using Lemma 3.10, for any $s \in [\sqrt{np}]$, we will show that there is a substantial number of rows of $\bar{A}_n$ which have one non-zero entry in the block J_s and no such entries in J'_s. Let us check it. On the event $\Omega_{\mathcal{L}}$,

$$|L_{\ell_\star}| \leq \frac{1}{c_{3.10}^\star p} \leq \frac{2n}{c_{3.10}^\star \log n} \leq n - n^{1/3} \leq n - |\mathcal{L}(\bar{A}_n)|,$$

where the second inequality uses assumption (3.19). Since $\hat{x}_{\mathcal{L}(\bar{A}_n)} = 0$, and $L_{\boldsymbol{\ell}_\star}$ contain the coordinates of $\hat{x}$ with the largest absolute value, it implies that $L_{\boldsymbol{\ell}_\star} \subset \mathcal{L}(\bar{A}_n)^c$, and hence $J_s \cap \mathcal{L}(\bar{A}_n) = \varnothing$. Moreover, $|J'_s| \leq |L_{\ell_\star}| = \kappa\sqrt{pn}$. Therefore we now apply Lemma 3.10 to get a set $\mathcal{A}$ such that $\mathcal{A}^c \cap \Omega_{3.7}$ has a small probability and on $\mathcal{A} \cap \Omega_{\mathcal{L}}$ there exist subsets of rows $I_s \subset [\bar{n}]$ with $|I_s| \geq c_{3.10}\kappa np$ for all $s \in [\sqrt{pn}]$, such that for every $i \in I_s$, we have $|\mathfrak{a}_{i,j_i}| = 1$ for only one index $j_i \in J_s$ and $\mathfrak{a}_{i,j} = 0$ for all $j \in J_s \cup J'_s\backslash\{j_0\}$. This means that $I_1, I_2, \ldots, I_{\sqrt{pn}}$ are disjoint subsets. Moreover $\{I_s\}_{s\in[\sqrt{np}]}$ satisfy the property (3.2). That is,

$$\operatorname{supp}(\operatorname{row}_i(\operatorname{fold}(\bar{A}_n))) \cap \mathcal{L}(\bar{A}_n) = \varnothing, \quad \text{for all } i \in I_s, \text{ and } s \in [\sqrt{np}]. \tag{3.20}$$

Therefore, for $s \in [\sqrt{np}]$ and $i \in I_s$,

$$\Big(\operatorname{fold}(\bar{A}_n) \cdot (x_{\mathcal{L}(\bar{A}_n)})\Big)_i = 0$$

and thus denoting $w := \frac{x_{\mathcal{L}(\bar{A}_n)}}{\|x_{[n]\backslash\mathcal{L}(\bar{A}_n)}\|_2}$ we deduce that

$$\big|\big(\operatorname{fold}(\bar{A}_n) \cdot (u+w)\big)_i\big| = (\operatorname{fold}(\bar{A}_n)u)_i| = |u_{i,j_i}| \geq |\hat{x}(\pi_{\hat{x}}^{-1}(s\kappa))|,$$

where the inequality follows from the monotonicity of the sequence $\{|\hat{x}(\pi_{\hat{x}}^{-1}(k))|\}_{k=1}^n$. Hence

$$\begin{aligned}\left\|\text{fold}(\bar{A}_n)\cdot(u+w)\right\|_2^2 \ge \sum_{s=1}^{\sqrt{np}}\sum_{i\in I_s}\left((\text{fold}(A_n)u)_i\right)^2 &\ge c_{3.10}np\sum_{s=1}^{(pn)^{1/2}}\kappa(\hat{x}(\pi_{\hat{x}}^{-1}(s\kappa)))^2\\ &\ge c_{3.10}np\sum_{k=(np)^{(\ell_\star-1)/2}}^{(np)^{\ell_\star/2}}(\hat{x}(\pi_{\hat{x}}^{-1}(k)))^2\\ &= c_{3.10}np\,\|\hat{z}_{\ell_\star}\|_2^2 \ge c_{3.10}np\cdot(c_{3.14}/K)^{2\ell_\star}. \end{aligned} \tag{3.21}$$

Combining this with the bound on $\|v\|_2$ (see (3.18)), we deduce that

$$\begin{aligned}\left\|\text{fold}(\bar{A}_n)\cdot\frac{x}{\|x_{[n]\setminus\mathcal{L}(\bar{A}_n)}\|_2}\right\|_2 &\ge \left\|\text{fold}(\bar{A}_n)(u+w)\right\|_2 - \left\|\text{fold}(\bar{A}_n)\right\|\cdot\|v\|_2\\ &\ge \sqrt{c_{3.10}np}\cdot(c_{3.14}/K)^{\ell_\star} - 4K\sqrt{np}\cdot 4(c_{3.14}/K)^{(\ell_\star+1)} \ge (\widetilde{c}_{3.14}/K)^{\ell_\star}\sqrt{np}, \end{aligned} \tag{3.22}$$

on the set $\mathcal{A}\cap\Omega_{\mathcal{L}}\cap\Omega_K\cap\Omega_{3.7}$, where the last inequality follows upon choosing $c_{3.14}$ and $\widetilde{c}_{3.14}$ sufficiently small (independently of $\ell_\star$).

Now it remains to consider the case $\ell_\star = \ell_0$. Note that in this case, using (3.21), we have that

$$\left\|\text{fold}(\bar{A}_n)\cdot(u+w)\right\|_2 \ge \sqrt{c_{3.10}np}\cdot\|\hat{z}_{\ell_0}\|_2\,,$$

and from (3.17), we have $\|v\|_2 = \|\hat{z}_{\ell_0+1}\|_2 \le 2c_{3.14}K^{-1}\|\hat{z}_{\ell_0}\|_2$. Therefore proceeding as before, on $\mathcal{A}\cap\Omega_K\cap\Omega_{\mathcal{L}}\cap\Omega_{3.7}$, we obtain

$$\left\|\text{fold}(\bar{A}_n)\cdot\frac{x}{\|x_{[n]\setminus\mathcal{L}(\bar{A}_n)}\|_2}\right\|_2 \ge (\widetilde{c}_{3.14}/K)^{\ell_\star}\sqrt{np}. \tag{3.23}$$

Since $np \le C_{3.13}\log n$, using Lemma 3.12 we also have that

$$\|x_{[n]\setminus\mathcal{L}(A_n)}\|_2 \ge \frac{1}{C_{3.12}np} \ge \frac{1}{C_{3.12}C_{3.13}\log n} \ge (c_{3.14}/K)^{\ell_\star},$$

on the set $\Omega_0^c\cap\Omega_{3.7}$, for all large n. Therefore, combining (3.22)-(3.23), and using Lemma 3.10 and Lemma 3.13 we establish (3.13) for all $p\in(0,1)$ such that $\log(1/p)\le np\le C_{3.13}\log n$.

Next we consider the case when $C_{3.13}\log n \le np \le (c_{3.10}^\star n)^{2/3}$. For such choices of $p\in(0,1)$ we use Lemma 3.13 to obtain that $\{\mathcal{L}(\bar{A}_n)=\varnothing\}$ with high probability. Using this fact one proceeds similarly as in the previous case to arrive at (3.13). Below is a brief outline.

Similarly to $\{\hat{z}_\ell\}_{\ell=1}^{\ell_0+1}$ defined in (3.15)-(3.16), we first define $\{z_\ell\}_{\ell=1}^{\ell_0+1}$ by rearranging the magnitudes of the coordinates of x and grouping them in blocks of length $(np)^{\ell/2}$ for $\ell = 1,2,\ldots,\ell_0$ and z_{ℓ_0+1} being the remaining block. Next, we define $\ell_\star$ to be the largest $\ell\le\ell_0$ such that $\|z_\ell\|_2 \ge (c_{3.14}/K)^\ell$. Equipped with the definition of $\ell_\star$ we then define $\{J_s,J_s'\}_{s\in[\sqrt{np}]}$ similarly as in the previous case. On the event $\{\mathcal{L}(\bar{A}_n)=\varnothing\}$ the requirement that $J_s\cap\mathcal{L}(\bar{A}_n)=\varnothing$ trivially follows. Since $np\ge C_{3.13}\log n$ by Lemma 3.13 we have that $\mathcal{L}(\bar{A}_n)=\varnothing$ with high probability. This allows us to use Lemma 3.10 to find disjoint subsets of rows $\{I_s\}_{s\in[\sqrt{np}]}$ with the desired properties and hence by repeating the same computations as in the previous case we arrive at (3.23). Now noting that $\|x_{[n]\setminus\mathcal{L}(\bar{A}_n)}\|_2 = 1$, on a set with high probability, we obtain the desired bound in (3.13).

It remains to provide a proof of (3.13) for $p\in(0,1)$ such that $np\ge(c_{3.10}^\star n)^{2/3}$. For this range of p we do not need the elaborate chaining argument of the previous two cases. It follows from the following simpler argument.

Fixing $x \in \text{Dom}\left(m, c_{3.14}K^{-1}\right)$, for $k \in \text{supp}(x_{[1:m]})$ we define $J_k = \{k\}$ and $J'_k = \text{supp}(x_{[1:m]})\backslash\{k\}$. Applying Lemma 3.10 with $\kappa = 1$ and $\mathfrak{m} = m$ we find disjoint subsets of rows $\{I_k\}_{k \in \text{supp}(x_{[1:m]})}$ such that $|I_k| \geq c_{3.10}np$ for all $k \in \text{supp}(x_{[1:m]})$ (note that by Lemma 3.13 $\mathcal{L}(\bar{A}_n) = \varnothing$ with high probability). Therefore, proceeding similarly as in (3.21) we obtain that

$$\left\| \text{fold}(\bar{A}_n) \cdot \frac{x_{[1:m]}}{\|x_{[1:m]}\|_2} \right\|_2 \geq \sum_{k \in \text{supp}(x_{[1:m]})} \sum_{i \in I_k} \frac{(\text{fold}(\bar{A}_n)x_{[1:m]})_i}{\|x_{[1:m]}\|_2} \geq c_{3.10}np \sum_{k \in \text{supp}(x_{[1:m]})} \frac{|x_k|^2}{\|x_{[1:m]}\|_2} = c_{3.10}np$$

on set $\mathcal{A} \cap \{\mathcal{L}(\bar{A}_n) = \varnothing\} \cap \Omega_{3.7}$ such that $\mathcal{A}^c \cap \Omega_{3.7}$ has a small probability. Using the fact that $x \in \text{Dom}\left(m, c_{3.14}K^{-1}\right)$ we observe that

$$\left\| \frac{x_{[1:m]}}{\|x_{[1:m]}\|_2} - x_{[1:m]} \right\|_2 = 1 - \|x_{[1:m]}\|_2 \leq \|x_{[m+1:n]}\|_2 \leq c_{3.14}K^{-1}\sqrt{m}\|x_{[m+1:n]}\|_\infty \leq c_{3.14}K^{-1}.$$

Thus applying the triangle inequality we deduce that for any $x \in \text{Dom}\left(m, c_{3.14}K^{-1}\right)$

$$\begin{aligned} \|\text{fold}(\bar{A}_n)x\|_2 &\geq \left\| \text{fold}(\bar{A}_n) \frac{x_{[1:m]}}{\|x_{[1:m]}\|_2} \right\|_2 - 4K\sqrt{np} \left(\left\| \frac{x_{[1:m]}}{\|x_{[1:m]}\|_2} - x_{[1:m]} \right\|_2 + \|x_{[m+1:n]}\|_2 \right) \\ &\geq \sqrt{c_{3.10}np} - 8c_{3.14}\sqrt{np} \geq \sqrt{\frac{c_{3.10}np}{2}}, \end{aligned}$$

on the event $\mathcal{A} \cap \Omega_K \cap \{\mathcal{L}(\bar{A}_n) = \varnothing\}$, whenever $c_{3.14} \leq \frac{1}{16}\sqrt{c_{3.10}}$. This together with Lemma 3.13 proves (3.13) for all $p \in (0, 1)$ such that $np \geq (c^\star_{3.10}n)^{2/3}$ and it finishes the proof of the proposition. $\square$

3.2. **Invertibility over vectors close to moderately sparse.** In this section we extend the uniform bound of Proposition 3.14 for vectors close to moderately sparse vectors. The following is the main result of this section.

Proposition 3.15. *Let A_n be as in Assumption 3.1, V_0, and ρ be as in Proposition 3.14, and $p \geq c_1 \frac{\log n}{n}$ for some constant $c_1 > 0$. Fix $K \geq 1$. Then there exist constants $0 < c_{3.15}, \widetilde{c}_{3.15}, c^*_{3.15}, \bar{c}_{3.15} < \infty$, depending only on K, such that for any M with $p^{-1} \leq M \leq c^*_{3.15}n$ and $y_0 \in \mathbb{R}^n$ we have*

$$\mathbb{P}\Big(\exists x \in \text{Dom}\left(M, c_{3.15}K^{-1}\right) \cup \text{Comp}(M, \rho)\backslash V_0 \text{ such that } \|A_nx - y_0\|_2 \leq \widetilde{c}_{3.15}\rho\sqrt{np}$$
$$\text{and } \|A_n - \mathbb{E}A_n\| \leq K\sqrt{np} \Big) \leq \exp(-\bar{c}_{3.15}n).$$

As outlined in Section 2 the key to the proof of Proposition 3.15 will be to obtain an estimate on the small ball probability. This will be achieved by deriving bounds on the Lévy concentration function (recall Definition 2.1). The necessary bound is derived in the lemma below.

Lemma 3.16. *Let C_n be a $n_1 \times n_2$ matrix, where $n_1, n_2 \geq \mathfrak{n}$ (recall $\mathfrak{n} := \lfloor n/2 \rfloor$), with i.i.d.* Ber$(p)$ *entries. Then for any $\alpha > 1$, there exist $\beta, \gamma > 0$, depending only on α such that for $x \in \mathbb{R}^{n_2}$, satisfying $\|x\|_\infty / \|x\|_2 \leq \alpha\sqrt{p}$, we have*

$$\mathcal{L}\left(C_nx, \beta \cdot \sqrt{np}\,\|x\|_2\right) \leq \exp(-\gamma n).$$

Lemma 3.16 is a consequence of [5, Corollary 3.7]. The difference between Lemma 3.16 and [5, Corollary 3.7] is that the latter has been proved for matrices whose entries have zero mean and obey a certain product structure. The key to the proof of [5, Corollary 3.7] is [5, Lemma 3.5]. Upon investigating the proof of [5, Lemma 3.5] it becomes evident that neither the zero mean condition nor the product structure of the entries are essential to its proof. So, repeating the proof of [5, Lemma 3.5] under the current set-up and following the proofs of [5, Lemma 3.6, Corollary 3.7] we derive Lemma 3.16. Further details are omitted.

We additionally borrow the following fact from the proof of [5, Lemma 3.8].

Fact 3.17. *Fix $M_1 < M_2 < n$ and for any $x \in S^{n-1}$ define*

$$u_x := u(x, M_1) := x_{[1:M_1]}, \quad v_x := v(x, M_1, M_2) := x_{[M_1+1:M_2]}, \text{ and } r_x := r(x, M_2) := x_{[M_2+1:n]}.$$

Then, given any $\varepsilon, \tau > 0$ and a set $\mathcal{S} \subset S^{n-1}$ there exists a set $\mathcal{M} \subset \mathcal{S}$ such that given any $x \in \mathcal{S}$ there exists a $\bar{x} \in \mathcal{M}$ such that

$$\|u_x - u_{\bar{x}}\|_2 \le \varepsilon, \quad \left\| \frac{v_x}{\|v_x\|_2} - \frac{v_{\bar{x}}}{\|v_{\bar{x}}\|_2} \right\|_2 \le \tau, \quad \text{and} \quad |\|v_x\|_2 - \|v_{\bar{x}}\|_2| \le \varepsilon. \tag{3.24}$$

and

$$|\mathcal{M}| \le \binom{n}{M_1}\binom{n-M_1}{M_2-M_1} \cdot \left(\frac{6}{\varepsilon}\right)^{M_1+1} \cdot \left(\frac{6}{\tau}\right)^{M_2-M_1}. \tag{3.25}$$

The proof of Fact 3.17 follows from volumetric estimates. Indeed, one first fixes the choice of the supports of u_x and v_x, and constructs standard nets for u_x, $v_x/\|v_x\|_2$, and $\|v_x\|_2$ of desired precision. Then bounds on the cardinality of follows by taking a union over the set of all possible choices of the supports of u_x and v_x. We omit further details.

Now we are ready to prove Proposition 3.15.

Proof of Proposition 3.15. Fix M with $p^{-1} \le M \le c^*_{3.15} n$ and for ease of writing let us denote

$$V_1 := \mathrm{Dom}\left(M, c_{3.15}K^{-1}\right) \cup \mathrm{Comp}(M, \rho), \tag{3.26}$$

where $c_{3.15}$ and $c^*_{3.15}$ to be determined during the course of the proof. We will show that for any $y \in \mathbb{R}^n$

$$\mathbb{P}\left(\{\exists x \in V_1 \backslash V_0 : \|(A_n - p\boldsymbol{J}_n)x - y\|_2 \le 2\widetilde{c}_{3.15}\rho\sqrt{np}\} \cap \Omega_K^0\right) \le \exp(-2\bar{c}_{3.15}n), \tag{3.27}$$

where $\boldsymbol{J}_n$ is the $n \times n$ matrix of all ones and

$$\Omega_K^0 := \{\|A_n - \mathbb{E}A_n\| \le K\sqrt{np}\}. \tag{3.28}$$

First let us show that the proposition follows from (3.27). To this end, denote

$$\mathcal{Y}_n := \{y \in \mathbb{R}^n : y = y_0 + \lambda\mathbf{1}, |\lambda| \le \sqrt{n}p\}.$$

It easily follows that $\mathcal{Y}_n$ has a net $\mathcal{Y}'_n$ of mesh size $\widetilde{c}_{3.15}\rho\sqrt{np}$ with cardinality at most $O(\sqrt{p}/\rho) = \exp(O(\log n))$. Therefore, noting that for any $y_1, y_2 \in \mathbb{R}^n$

$$\left| \inf_{x \in V_1 \backslash V_0} \|(A_n - p\boldsymbol{J}_n)x - y_1\|_2 - \inf_{x \in V_1 \backslash V_0} \|(A_n - p\boldsymbol{J}_n)x - y_2\|_2 \right| \le \|y_1 - y_2\|_2,$$

taking a union over $y \in \mathcal{Y}'_n$ we deduce from (3.27) that

$$\mathbb{P}\left(\left\{\inf_{x \in V_1 \backslash V_0, y \in \mathcal{Y}_n} \|(A_n - \boldsymbol{J}_n p)x - y\|_2 \le \widetilde{c}_{3.15}\rho\sqrt{np}\right\} \cap \Omega_K^0\right) \le \exp(-\bar{c}_{3.15}n). \tag{3.29}$$

Since $y_0 - \boldsymbol{J}_n x \in \mathcal{Y}_n$ for all $x \in S^{n-1}$ we further note that

$$\inf_{x \in V_1 \backslash V_0, y \in \mathcal{Y}_n} \|(A_n - \boldsymbol{J}_n p)x - y\|_2 \le \inf_{x \in V_1 \backslash V_0} \|A_n x - y_0\|_2. \tag{3.30}$$

This together with (3.29) yields the desired conclusion. Thus, it remains to establish (3.27).

To this end, we fix any $x \in V_1 \backslash V_0, y \in \mathbb{R}^n$ and write

$$A_n := \begin{bmatrix} A_n^{1,1} & A_n^{1,2} \\ A_n^{2,1} & A_n^{2,2} \end{bmatrix}, \quad x := \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \text{and} \quad y := \begin{pmatrix} y_1 \\ y_2 \end{pmatrix},$$

where $A_n^{1,1}$ and $A_n^{2,2}$ are $\mathfrak{n} \times \mathfrak{n}$ and $(n - \mathfrak{n}) \times (n - \mathfrak{n})$ matrices, respectively, $A_n^{1,2}$, and $(A_n^{2,1})^*$ are $\mathfrak{n} \times (n - \mathfrak{n})$ matrices, x_1, y_1 are vectors of length $\mathfrak{n}$, and x_2, y_2 are vectors of length $(n - \mathfrak{n})$. Similarly we define $\{\boldsymbol{J}_n^{i,j}\}_{i,j=1}^2$. With these notations we see that

$$\|(A_n - p\boldsymbol{J}_n)x - y\|_2^2 = \|(A_n^{1,1} - p\boldsymbol{J}_n^{1,1})x_1 + (A_n^{1,2} - p\boldsymbol{J}_n^{1,2})x_2 - y_1\|_2^2 + \|(A_n^{2,1} - p\boldsymbol{J}_n^{2,1})x_1 + (A_n^{2,2} - p\boldsymbol{J}_n^{2,2})x_2 - y_2\|_2^2.$$

Also note that by Assumption 3.1 both $A_n^{1,2}$ and $A_n^{2,1}$ are matrices with i.i.d. Bernoulli entries independent of $A_n^{1,1}$ and $A_n^{2,2}$, respectively. Since $x \notin V_0$ we have $\|x_{[m+1:n]}\|_\infty / \|x_{[m+1:n]}\|_2 \le c_{3.14}^{-1} K m^{-1/2}$. Further, for $j \in [n]$, let us define

$$x_1'(j) := x_{[m+1:n]}(j) \cdot \mathbb{I}(j \in [\mathfrak{n}]), \quad \text{and } x_2'(j) := x_{[m+1:n]}(j) \cdot \mathbb{I}(j \notin [\mathfrak{n}]).$$

Therefore there exists $i \in \{1, 2\}$ such that $\|x_i'\|_2 \ge \|x_{[m+1:n]}\|_2 / \sqrt{2}$. Without loss of generality let us assume $i = 1$. This implies that $\|x_1'\|_\infty / \|x_1'\|_2 \le 2c_{3.14}^{-1} K m^{-1/2} = 2c_{3.14}^{-1} (c_{3.10}^\star)^{-1/2} K \sqrt{p}$. Hence applying Lemma 3.16 we see that for a sufficiently small $\widetilde{c}_{3.15}$, we have

(3.31)
$$\mathbb{P}\left(\|(A_n - p\boldsymbol{J}_n)x - y\|_2 \le 4\widetilde{c}_{3.15} \|x_{[m+1:n]}\|_2 \sqrt{np}\right) \le \mathcal{L}\left(A_n^{2,1} x_1', 8\widetilde{c}_{3.15} \|x_1'\|_2 \sqrt{np}\right) \le \exp(-3\bar{c}n),$$

for some $\bar{c} > 0$.

To finish the proof we now use a ε-net argument. Applying Fact 3.17 we see that there exists $\mathcal{M} \subset V_1 \backslash V_0$ such that for any $x \in V_1 \backslash V_0$, there exists $\bar{x} \in \mathcal{M}$ so that (3.24) holds. Thus

$$\|v_x - v_{\bar{x}}\|_2 \le \left\| \frac{v_x}{\|v_x\|_2} - \frac{v_{\bar{x}}}{\|v_{\bar{x}}\|_2} \right\|_2 \|v_{\bar{x}}\|_2 + \|v_x\|_2 \left| 1 - \frac{\|v_{\bar{x}}\|_2}{\|v_x\|_2} \right| \le \varepsilon + \tau \|v_{\bar{x}}\|_2. \tag{3.32}$$

Since $\bar{x} \in V_1$ we also observe that

$$\|r_{\bar{x}}\|_2 \le c_{3.15} K^{-1} \sqrt{M} \|r_{\bar{x}}\|_\infty \le 2c_{3.15} K^{-1} \|v_{\bar{x}}\|_2, \tag{3.33}$$

where the last inequality follows from the facts that the coordinates of $r_{\bar{x}}$ have smaller magnitudes than the non-zero coordinates of $v_{\bar{x}}$ and $m \le M/2$. Since $\bar{x} \notin V_0$, we have

$$\|\bar{x}_{[m+1:n]}\|_2 = \sqrt{\|v_{\bar{x}}\|_2^2 + \|r_{\bar{x}}\|_2^2} \ge \rho. \tag{3.34}$$

Therefore, it follows from above that $\|v_{\bar{x}}\|_2 \ge \|r_{\bar{x}}\|_2$, whenever $c_{3.15}$ chosen sufficiently small, which further implies that $\|v_{\bar{x}}\|_2 \ge \rho/\sqrt{2}$. Hence, choosing $\varepsilon \le \rho/\sqrt{2}$, and proceeding similarly as in (3.33) we deduce

$$\|r_x\|_2 \le 2c_{3.15} K^{-1} (\|v_{\bar{x}}\|_2 + \varepsilon) \le 4c_{3.15} K^{-1} \|v_{\bar{x}}\|_2. \tag{3.35}$$

Further note that

$$\|A_n - p\boldsymbol{J}_n\| \le \|A_n - \mathbb{E}A_n\| + \|\mathbb{E}A_n - p\boldsymbol{J}_n\| \le \|A_n - \mathbb{E}A_n\| + 1,$$

where the last step follows from Assumption 3.1. So, using the triangle inequality, (3.24), (3.32)-(3.33), and (3.35) we deduce

$$\begin{aligned}\|(A_n - p\boldsymbol{J}_n)\bar{x} - y\|_2 &\le \|(A_n - p\boldsymbol{J}_n)x - y\|_2 \\ &\quad + \|A_n - p\boldsymbol{J}_n\| \Big(\|u_x - u_{\bar{x}}\|_2 + \|v_x - v_{\bar{x}}\|_2 + \|r_x\|_2 + \|r_{\bar{x}}\|_2 \Big) \\ &\le \|(A_n - p\boldsymbol{J}_n)x - y\|_2 + 4K\sqrt{np}\cdot\varepsilon + 2K\sqrt{np}\cdot\tau\cdot\|v_{\bar{x}}\|_2 + 12c_{3.15}\sqrt{np}\cdot\|v_{\bar{x}}\|_2 . \end{aligned} \tag{3.36}$$

Thus setting

$$\varepsilon = \frac{c_{3.15}\rho}{4K} \quad \text{and} \quad \tau = \frac{c_{3.15}}{2K}, \tag{3.37}$$

and shrinking $c_{3.15}$ further, from (3.31) and (3.34) we derive that

$$\begin{aligned}&\mathbb{P}\left(\exists x \in V_1\backslash V_0 : \|(A_n - p\boldsymbol{J}_n)x - y\|_2 \le 2\widetilde{c}_{3.15}\rho\sqrt{np}\right) \\ &\le \mathbb{P}\left(\exists \bar{x} \in \mathcal{M} : \|(A_n - p\boldsymbol{J}_n)\bar{x} - y\|_2 \le 4\widetilde{c}_{3.15}\|\bar{x}_{[m+1:n]}\|_2\sqrt{np}\right) \le |\mathcal{M}|\cdot\exp(-3\bar{c}n). \end{aligned} \tag{3.38}$$

With the above choices of ε and τ, and any $M \le c^*_{3.15}n$, from (3.25) we have that

$$\begin{aligned}|\mathcal{M}| \le \bar{C}^{2M}\binom{n}{m}\binom{n}{M}\cdot\left(\frac{1}{\rho}\right)^{m+1} &\le \bar{C}^{2c^*_{3.15}n}\left(\frac{en}{m}\right)^m\left(\frac{en}{M}\right)^M\left(\frac{1}{\rho}\right)^{m+1} \\ &\le (e(c^\star_{3.10})^{-1}np)^m\left(\bar{C}^2\frac{e}{c^*_{3.15}}\right)^{c^*_{3.15}n}\left(\frac{1}{\rho}\right)^{m+1},\end{aligned}$$

for some constant $\bar{C}$ depending only on K. Recalling the definition of ρ and m, it is easy to note that

$$m\log\left(\frac{np}{\rho}\right) = o(n),$$

for all p satisfying $np \ge \frac{\log n}{\sqrt{\log\log n}}$. This implies that for $c^*_{3.15}$ sufficiently small, if $M \le c^*_{3.15}n$ then we have $|\mathcal{M}| \le \exp(\bar{c}n)$. In combination with (3.38), this yields (3.27). The proof of the proposition is complete. ☐

3.3. **Invertibility over sparse vectors with a large spread component.** Combining Proposition 3.14 and Proposition 3.15 we see that we have a uniform lower bound on $\|A_nx\|_2$ for $x \in V_1$ (recall the definition V_1 from (3.26)) with $M = c^*_{3.15}n$. As seen from the proof of Proposition 3.15, the positive constant $c^*_{3.15}$ is small. On the other hand, as we will see in Section 4, to obtain a uniform lower bound on $\|A_nx\|_2$ over incompressible and non-dominated vectors x in the case when A_n is the adjacency matrix of a directed Erdős-Rényi graph, we first need to prove a uniform lower bound on the same for $x \in V_{c^*,c}$, where

$$V_{c^*,c} := \mathrm{Dom}(c^*n, cK^{-1}) \cup \mathrm{Comp}(c^*n, \rho), \tag{3.39}$$

with the constant c^* close to one (in fact $c^* > \frac{3}{4}$ will do) and $c > 0$ some another constant. This is not immediate from Proposition 3.15 and it will be the main result of this short section.

Proposition 3.18. *Let A_n be as in Assumption 3.1, ρ as in Proposition 3.14, and $p \ge c_1\frac{\log n}{n}$ for some constant $c_1 \in (0,1)$. Fix $K \ge 1$ and $c^*_0 \in (0,1)$. Let $M_0 := \frac{n\sqrt{\log\log n}}{\log n}$. Then there exist constants $0 < c_{3.18}, \widetilde{c}_{3.18}, \bar{c}_{3.18} < \infty$, depending only on c^*_0 and K, such that for any $y \in \mathbb{R}^n$ we have*

$$\mathbb{P}\left(\left\{\exists x \in V_{c^*_0,c_{3.18}}\backslash V_{M_0} : \|(A_n - p\boldsymbol{J}_n)x - y\|_2 \le 4\widetilde{c}_{3.18}\|x_{[M_0+1:c^*_0n]}\|_2\sqrt{np}\right\} \cap \Omega^0_K\right) \le \exp(-2\bar{c}_{3.18}n), \tag{3.40}$$

for all large n, where

$$V_{M_0} := \mathrm{Dom}(M_0, c_{3.15}K^{-1}) \cup \mathrm{Comp}(M_0, \rho). \tag{3.41}$$

Consequently, for any $y_0 \in \mathbb{R}^n$ we have

$$\mathbb{P}\Big(\exists x \in V_{c_0^*, c_{3.18}} \backslash V_{M_0} \text{ such that } \|A_n x - y_0\|_2 \le \widetilde{c}_{3.18}\rho\sqrt{np} \text{ and } \|A_n - \mathbb{E}A_n\| \le K\sqrt{np}\Big) \le \exp(-\bar{c}_{3.18} n). \tag{3.42}$$

As the set of sparse vectors that have a large spread component has a higher metric entropy compared to that of the set of vectors considered in Section 3.2, the small ball probability estimate derived in Lemma 3.16 will be insufficient to accommodate a union bound. To obtain a useful bound on the small ball probability we use the following result. Before stating the lemma let us introduce a notation: for any $v \in \mathbb{R}^n$ and $J \subset [n]$ we write v_J to denote the vector in $\mathbb{R}^n$ obtained from v by setting $v_i = 0$ for all $i \in J^c$.

Lemma 3.19 (Bound on Lévy concentration function)**.** *Let $v \in \mathbb{R}^n$ be a fixed vector and $\boldsymbol{x} \in \mathbb{R}^n$ be a random vector with i.i.d.* $\mathrm{Ber}(p)$ *for some $p \in (0,1)$. Then there exists an absolute constant $C_{3.19}$ such that for every $\varepsilon > 0$ and $J \in [n]$,*

$$\mathcal{L}\left(\langle \boldsymbol{x}, v\rangle, p^{1/2}(1-p)^{1/2}\|v_J\|_2\varepsilon\right) \le \mathcal{L}\left(\langle \boldsymbol{x}_J, v_J\rangle, p^{1/2}(1-p)^{1/2}\|v_J\|_2\varepsilon\right) \le C_{3.19}\left(\varepsilon + \frac{\|v_J\|_\infty}{p^{1/2}(1-p)^{1/2}\|v_J\|_2}\right).$$

The proof of Lemma 3.19 is a simple consequence of the well known Berry-Esséen theorem and is similar to that of [31, Proposition 3.2]. Hence further details are omitted.

To utilize the bound from Lemma 3.19 we recall that any vector belonging to the third set has a large spread component. This means that one can find a $J \subset [n]$ such that $\|v_J\|_\infty / \|v_J\|_2$ is small with the Euclidean norm of v_J being not too small.

The proof of Proposition 3.18 is similar to that of Proposition 3.15. Recall a key to the proof of Proposition 3.15 is the anti-concentration bound of Lemma 3.16 where the latter is a consequence of Paley-Zygmund inequality (see the proof of [5, Corollary 3.7]). To prove Proposition 3.18 we need a better anti-concentration bound. To this end, we note that any $x \notin V_{M_0}$ has a large spread component, i.e. a large non-dominated part. It allows us to use Lemma 3.19 instead of Paley-Zygmund inequality. For matrices with independent rows, this together with standard tensorization techniques produces a sharp enough anti-concentration probability bound suitable for the proof of Proposition 3.18. For matrices satisfying Assumption 3.1 we additionally need to show that one can find a sub-matrix of A_n with jointly independent entries, such that the coordinates of x which correspond to the columns of this sub-matrix form a vector with a large spread component and a sufficiently large norm to carry out the scheme described above. Since the proof of Proposition 3.18 is an adaptation of that of Proposition 3.15 with these couple of modifications it is deferred to Appendix B.

Remark 3.20. Proposition 3.18 shows that $\|A_n x\|_2$ has uniform lower bound when $x \in V_{c_0^*, c_{3.18}} \backslash V_{M_0}$. Its proof reveals that the same bound continues to hold when V_{M_0} is replaced by

$$\bar{V}_{M_0} := \mathrm{Dom}(M_0, cK^{-1}) \cup \mathrm{Comp}(M_0, \rho),$$

for some small constant $c < c_{3.15}$. Changing the constant $c_{3.15}$ to c only shrinks the constants $c_{3.18}, \widetilde{c}_{3.18}, \bar{c}_{3.18}$. We will use this generalization in the proof of Lemma 3.23.

Remark 3.21. Propositions 3.14, 3.15, and 3.18 have been proved for $n \times n$ matrices. It can be checked that the conclusions of these propositions continue to hold for $(n-1) \times n$ matrices, with slightly worse constants. In particular, they hold for the matrix $\widetilde{A}_n$ such that its rows are any $(n-1)$ columns of the matrix A_n satisfying Assumption 3.1. We will need this generalization to prove the desired lower bound on the smallest singular value of the adjacency matrix of a random bipartite graph or equivalently for the random matrix with i.i.d. Bernoulli entries (as noted in Remark 1.8). To keep the presentation of this paper simple we refrain from providing the proof for this generalization. It follows from a simple adaptation of the proof of the same for square matrices.

3.4. **Structure of $A_n^{-1}u$.** As mentioned in Section 2, to deduce invertibility over non-dominated and incompressible vectors we also need to show that, given any $u \in \mathbb{R}^n$, the random vector $A_n^{-1}u$ must be non-dominated and incompressible with high probability. Since we will apply this result with coordinates of u being i.i.d. $\text{Ber}(p)$, we may and will assume that u does not have a large support. With some additional work, the results of Sections 3.1-3.3 yield this.

Moreover, as we will see in Section 4, to treat the non-dominated and incompressible vectors when A_n is the adjacency matrix of a directed Erdős-Rényi graph, we further need to establish that given any $J \subset [n]$ with $|J| \approx \frac{n}{2}$, one can find $I \subset J$ such that the vector $(A_n^{-1}u)_I$ contains a considerable proportion the non-dominated and incompressible components of the random vector $A_n^{-1}u$. The proof of the latter crucially uses Proposition 3.18. These two results are the content of this section.

We first begin with the corollary which shows that $A_n^{-1}u$ is neither compressible nor dominated with high probability.

Corollary 3.22. *Let A_n be as in Assumption 3.1, where p satisfies the inequality*

$$np \geq \log(1/p),$$

and ρ be as in Proposition 3.14. Fix $K \geq 1$, $c_0^* \in (0,1)$, and $y_0 \in \mathbb{R}^n$ such that $|\text{supp}(y_0)| \leq C_\star np$ for some $C_\star > 0$. Then there exist constants $0 < \widetilde{c}_{3.22}, \bar{c}_{3.22} < \infty$, depending only on c_0^* and K, such that

$$\mathbb{P}\Big(\Big\{\exists x \in \mathbb{R}^n \text{ such that } x/\|x\|_2 \in V_{c_0^*, c_{3.18}}, \, \|A_n x - y_0\|_2 \leq \widetilde{c}_{3.22}\rho\sqrt{np}\cdot\|x\|_2 \text{ and } \|A_n - \mathbb{E}A_n\| \leq K\sqrt{np}\Big\} \cap \Omega_0^c\Big) \leq n^{-\bar{c}_{3.22}},$$

where we recall the definition of $V_{c_0^*, c_{3.18}}$ from (3.39) and the definition of Ω_0 from (1.1).

Proof. Recalling the definition of V_0 from (3.10), we first show that

$$\mathbb{P}\left(\{\exists x \in \mathbb{R}^n \text{ such that } x/\|x\|_2 \in V_0, \, \|A_n x - y_0\|_2 \leq \widetilde{c}_{3.22}\rho\sqrt{np}\cdot\|x\|_2\} \cap \Omega_K^0 \cap \Omega_0^c\right) \leq n^{-2\bar{c}_{3.22}}, \tag{3.43}$$

for all large n, where we recall the definition of Ω_K^0 from (3.28). We remind the reader that to prove Proposition 3.14 we defined $\bar{A}_n$ to be the sub-matrix of A_n obtained upon deleting the rows in $\text{supp}(y_0)$ and showed that $\|\bar{A}_n x\|_2$ is uniformly bounded below, with high probability, for all $x \in V_0$. As $\|\bar{A}_n x\|_2 \leq \|A_n x - y_0\|_2$ this yielded the desired result. Since the proof does not involve y_0, except for the cardinality of its support, we therefore can carry out the exact same steps and use the bound on the probability of $\Omega_{3.7}^c$, derived in Lemma 3.7, to obtain (3.43).

It remains to show that

$$\mathbb{P}\left(\left\{\exists x \in \mathbb{R}^n \text{ such that } x/\|x\|_2 \in V_{c_0^*,c_{3.18}}\backslash V_0,\ \|A_n x - y_0\|_2 \le \widetilde{c}_{3.22}\rho\sqrt{np}\cdot\|x\|_2\right\} \cap \Omega_K^0\right) \le \exp(-\bar{c}n), \tag{3.44}$$

for some $\bar{c} > 0$. If there exists an $x \in \mathbb{R}^n$ such that $\|A_n x - y_0\| \le \widetilde{c}_{3.22}\rho\sqrt{np}\cdot\|x\|_2$, then using triangle inequality we find that

$$\frac{\|y_0\|_2}{\|x\|_2} \le \|A_n - p\boldsymbol{J}_n\| + p\|\boldsymbol{J}_n\| + \frac{\|A_n x - y_0\|_2}{\|x\|_2} \le 2Knp,$$

on Ω_K^0. Further let us recall that for any $x \in S^{n-1}$, we have $p\boldsymbol{J}_n x = \lambda\mathbf{1}$ for some $\lambda \in \mathbb{R}$ with $|\lambda| \le \sqrt{np}$. Therefore, using triangle inequality once more we see that

$$\left\{\exists x \in \mathbb{R}^n \text{ such that } x/\|x\|_2 \in V_{c_0^*,c_{3.18}}\backslash V_0,\ \|A_n x - y_0\|_2 \le \widetilde{c}_{3.22}\rho\sqrt{np}\cdot\|x\|_2\right\} \cap \Omega_K^0$$
$$\subset \left\{\inf_{y\in\mathcal{Y}_\star}\ \inf_{x \in V_{c_0^*,c_{3.18}}\backslash V_0} \|(A_n - p\boldsymbol{J}_n)x - y\|_2 \le \widetilde{c}_{3.22}\rho\sqrt{np}\right\} \cap \Omega_K^0,$$

where

$$\mathcal{Y}_\star := \left\{\gamma\cdot\frac{y_0}{\|y_0\|_2} + \lambda\mathbf{1};\ \gamma,\lambda \in \mathbb{R} \text{ with } |\gamma| \le 2Knp, |\lambda| \le \sqrt{np}\right\}.$$

Since $\mathcal{Y}_\star$ admits a net $\mathcal{N}_\star$ of mesh size $\widetilde{c}_{3.22}\rho\sqrt{np}$ with cardinality at most $(8K\sqrt{np}/\widetilde{c}_{3.22}\rho)^2$, by a union bound we see that it suffices to show that

$$\mathbb{P}\left(\left\{\exists x \in V_{c_0^*,c_{3.18}}\backslash V_0, \text{ such that } \|(A_n - p\boldsymbol{J}_n)x - y\|_2 \le 2\widetilde{c}_{3.22}\rho\sqrt{np}\right\} \cap \Omega_K^0\right) \le \exp(-2\bar{c}n), \tag{3.45}$$

for any $y \in \mathbb{R}^n$. Arguing similarly as in (3.33), we note that for any $x \in \text{Dom}(c_0^* n, c_{3.18}K^{-1})$ $\|x_{[M_0+1:c_0^*n]}\|_2 \ge \|x_{[c_0^*n+1:n]}\|_2$, and hence, for $x \in \text{Dom}(c_0^*n, c_{3.18}K^{-1})\backslash V_{M_0}$ we obtain that $\|x_{[M_0+1:c_0^*n]}\|_2 \ge \rho/\sqrt{2}$. Therefore, (3.45) follows from (3.27) and (3.40). This yields (3.44) and combining this with (3.43) now finishes the proof of the corollary. ☐

Building on Corollary 3.22 we now prove that for any $J \subset [n]$ with $|J| \approx \frac{n}{2}$, there exists a large set $I \subset J$ such that $(A_n^{-1}u)_I$ has non-dominated tails and a substantial Euclidean norm.

Lemma 3.23. *Let A_n be as in Assumption 3.1, where p satisfies the inequality*

$$np \ge \log(1/p).$$

Fix $K \ge 1$, $J \subset [n]$ such that $\frac{3n}{8} \le |J| \le \frac{5n}{8}$, and $y_0 \in \mathbb{R}^n$ with $\|y_0\|_2 \in [1, \bar{C}np]$ and $|\text{supp}(y_0)| \le C_\star np$ for some $\bar{C}, C_\star > 0$. Then there exist constants $0 < c_{3.23}, \bar{c}_{3.23} < \infty$, depending only on K, such that

$$\mathbb{P}\left(\left\{\exists x \in \mathbb{R}^n : A_n x = y_0 \text{ and either } \frac{\|x_{[\frac{n}{4}+1:n]\cap J}\|_\infty}{\|x_{[\frac{n}{4}+1:n]\cap J}\|_2} \ge \frac{1}{c_{3.23}\sqrt{n}} \text{ or } \|x_{[\frac{n}{4}+1:n]\cap J}\|_2 \le \rho\right\} \cap \Omega_K^0 \cap \Omega_0^c\right) \le n^{-\bar{c}_{3.23}},$$

for all large n.

Proof. Let $x \in \mathbb{R}^n$ be such that $A_n x = y_0$. Let us show first that the event $\|x_{[\frac{n}{4}+1:n]\cap J}\|_2 \le \rho$ can occur with probability at most $n^{-\bar{c}}$ for some constant $\bar{c} > 0$. Since $|J| \ge \frac{3n}{8}$, we have

$$\|x_{J\cap[\frac{n}{4}+1:n]}\|_2 \ge \|x_{[\frac{7}{8}n+1:n]}\|_2,$$

where by a slight abuse of notation, for $m < m' < n$, we write

$$x_{J\cap[m:m']}(i) = x_{[m;m']}(i)\cdot \mathbf{1}(i\in J).$$

Hence, our claim follows by applying Corollary 3.22 with $c_0^* = \frac{7}{8}$.

Next, assume that

$$\left\| x_{J\cap[\frac{n}{4}+1:n]}\right\|_2 \le c_{3.23}\sqrt{n}\left\| x_{J\cap[\frac{n}{4}+1:n]}\right\|_\infty. \tag{3.46}$$

We will prove that if $c_{3.23}$ is chosen sufficiently small then this can hold only on a set of small probability as well. This will complete the proof.

Denote $w := x_{J^c\cup[1:\frac{n}{4}]}$ and $z := x - w = x_{J\cap[\frac{n}{4}+1:n]}$. Then $w\neq 0$ and the assumption $|J|\ge \frac{3n}{8}$ implies that $w\in \text{Sparse}(7n/8)$. We will show that the vector $A_n w/\|w\|_2$ is close to some set $\mathcal{Y}$ having a small ε-net. As $w/\|w\|_2\in \text{Sparse}(7n/8)\cap S^{n-1}$, the desired probability estimate will then follow from Proposition 3.18 and the union bound over the net.

Turning to carry out the above task, we note that the inequality (3.46) shows that

$$\|z\|_2 = \left\| x_{J\cap[\frac{n}{4}+1:n]}\right\|_2 \le 3c_{3.23}\left\| x_{[M_0+1:\frac{n}{4}]}\right\|_2,$$

where we recall $M_0 := \frac{n\sqrt{\log\log n}}{\log n}$. Since $A_n x = y_0$, this implies

$$\|A_n w - y_0 + p\boldsymbol{J}_n z\|_2 = \|(A_n - p\boldsymbol{J}_n)z\|_2 \le \|A_n - p\boldsymbol{J}_n\|\cdot\|z\|_2 \le 6c_{3.23}K\sqrt{np}\cdot\left\| x_{[M_0+1:\frac{n}{4}]}\right\|_2,$$

on the event Ω_K^0, where we recall its definition from (3.28).

For ease of writing let us denote $w^\star := w/\|w\|_2$ and set $y^\star := (y_0 - p\boldsymbol{J}_n z)/\|w\|_2$. With this notation, the previous inequality reads

$$\|A_n w^\star - y^\star\|_2 \le 6c_{3.23}K\sqrt{np}\cdot\frac{\|x_{[M_0+1:\frac{n}{4}]}\|_2}{\|w\|_2} = 6c_{3.23}K\sqrt{np}\cdot\left\| w^\star_{[M_0+1:\frac{n}{4}]}\right\|_2, \tag{3.47}$$

where we used that $w_{[M_0+1:n/4]} = x_{[M_0+1:n/4]}$ to derive the last equality.

The inequality (3.47) already shows that $A_n w/\|w\|_2$ is close to $y^\star$. From the definition of $y^\star$ we further note that $y^\star = \lambda y_0 + \gamma\mathbf{1}$ for some $\lambda,\gamma\in\mathbb{R}$. This indicates that the natural choice for the set $\mathcal{Y}$ is the collection of all vectors of the form $\lambda y_0 + \gamma\mathbf{1}, \lambda,\gamma\in\mathbb{R}$. To show that $\mathcal{Y}$ admits a net of small cardinality we need bounds on λ and γ.

We claim that $y^\star\in\mathcal{Y}$, where

$$\mathcal{Y} := \{y\in\mathbb{R}^n : y = \lambda y_0 + \gamma\mathbf{1}, \text{ for some } \lambda\in(0,4Knp] \text{ and } \gamma\in[-3\sqrt{np}, 3\sqrt{np}]\}.$$

To see this we observe that the assumption $A_n x = y_0$ implies that

$$\|x\|_2 \ge \frac{\|A_n x\|_2}{\|A_n\|} \ge \frac{\|y_0\|_2}{2Knp} \ge \frac{1}{2Knp},$$

as $\|y_0\|_2\ge 1$ and $\|A_n\|\le 2Knp$ on the event Ω_K^0. Therefore,

$$\|w\|_2 \ge \|x_{[1:\frac{n}{4}]}\|_2 \ge \frac{1}{2}\|x\|_2 \ge 1/(4Knp).$$

From this and the inequality $\|z\|_2\le 2\|w\|_2$ the required claim follows.

Since $\|y_0\|_2\le \bar{C}np$ it is also immediate that the set $\mathcal{Y}$ admits a $(\widetilde{c}_{3.18}\rho\sqrt{np})$-net $\mathcal{N}$ of cardinality at most $O(((np)^2/\rho)^2)$.

We next claim that $w^\star\notin\widetilde{V}_{M_0}$, with high probability, where

$$\widetilde{V}_{M_0} := \text{Dom}(M_0, \frac{c_{3.15}}{3}K^{-1}) \cup \text{Comp}(M_0,\rho). \tag{3.48}$$

Proving (3.48) will put us in a position to apply Proposition 3.18.

To this end, using Corollary 3.22, we can assume that $x/\|x\|_2 \notin V_{15/16,c_{3.18}}$, with high probability. Hence, recalling the definition of w and using the monotonicity of the non-zero coordinates of $x_{[M_0+1:n]}$ we have

$$\frac{\|w_{[M_0+1:n]}\|_2}{\|w\|_2} \geq \frac{1}{2} \cdot \frac{\|x_{[n/8+1:n/4]}\|_2}{\|x\|_2} \geq \frac{\|x_{[15n/16+1:n]}\|_2}{\|x\|_2} \geq \rho. \tag{3.49}$$

Moreover, $w_{[M_0+1:n/4]} = x_{[M_0+1:n/4]}$, and $\|x_{[M_0+1:n/4]}\|_2 \geq \frac{1}{3}\|x_{[M_0+1:n]}\|_2$, so

$$\frac{\|w_{[M_0+1:n]}\|_\infty}{\|w_{[M_0+1:n]}\|_2} \leq \frac{\|x_{[M_0+1:n]}\|_\infty}{\|x_{[M_0+1:n/4]}\|_2} \leq 3 \cdot \frac{\|x_{[M_0+1:n]}\|_\infty}{\|x_{[M_0+1:n]}\|_2} \leq \frac{3K}{c_{3.15}\sqrt{M_0}}, \tag{3.50}$$

where we applied Corollary 3.22 again in assuming that $x/\|x\|_2 \notin V_{M_0}$, with high probability, with V_{M_0} as in (3.41). Inequalities (3.49) and (3.50) confirm that $w^\star \notin \widetilde{V}_{M_0}$ with high probability.

Now, setting $c_0^* = \frac{7}{8}$ in Proposition 3.18 (see also Remark 3.20), combining (3.40) with the union bound over the net $\mathcal{N} \subset \mathcal{Y}$, and applying triangle inequality we derive that

$$\mathbb{P}\left(\left\{\inf_{y\in\mathcal{Y}} \inf_{w^\star \in \text{Sparse}(7n/8)\cap S^{n-1}\setminus \widetilde{V}_{M_0}} \|(A_n - p\boldsymbol{J}_n)w_\star - y\|_2 \leq 4\widetilde{c}_{3.18}\|w^\star_{[M_0+1:\frac{7n}{8}]}\|_2\sqrt{np}\right\} \cap \Omega_K^0\right) \leq \exp(-\bar{c}_{3.18}n).$$

Thus recalling (3.47), and as $y^\star \in \mathcal{Y}$, we see that for a sufficiently small $c_{3.23}$, the inequality (3.46) can hold only on a set of small probability. This finishes the proof of the lemma. □

4. Invertibility over incompressible and non-dominated vectors

In this section our goal is to obtain a uniform lower bound on $\|A_n x\|_2$ over non-dominated and incompressible vectors x, with large probability. As the set of such vectors possesses a large metric entropy, one cannot replicate the approach of Section 3. As outlined in Section 2, we find a uniform lower bound over the set of such vectors by relating it to the average of the distance of a column of A_n from the subspace spanned by the rest of the columns. To this end, we use the following Lemma from [36] (see Lemma 3.5 there).

Lemma 4.1 (Invertibility via distance)**.** *For $j \in [n]$, let $\tilde{A}_{n,j} \in \mathbb{R}^n$ be the j-th column of $\tilde{A}_n$, and let $\tilde{H}_{n,j}$ be the subspace of $\mathbb{R}^n$ spanned by $\{\tilde{A}_{n,i}, i \in [n]\setminus\{j\}\}$. Then for any $\varepsilon, \rho > 0$, and $M < n$,*

$$\mathbb{P}\left(\inf_{x\in\text{Incomp}(M,\rho)} \left\|\tilde{A}_n x\right\|_2 \leq \varepsilon\rho^3\sqrt{\frac{p}{n}}\right) \leq \frac{1}{M}\sum_{j=1}^n \mathbb{P}\left(\text{dist}(\tilde{A}_{n,j}, \tilde{H}_{n,j}) \leq \rho^2\sqrt{p}\varepsilon\right). \tag{4.1}$$

Remark 4.2. Lemma 4.1 can be extended to the case when the event in the LHS of (4.1) is intersected with an event Ω. In that case Lemma 4.1 continues to hold if the RHS of (4.1) is replaced by intersecting each of the event under the summation sign with the same event Ω. In the proof of Theorem 1.7, we will use this slightly more general version of Lemma 4.1. Since the proof of this general version of Lemma 4.1 is a straightforward adaptation of the proof of [36, Lemma 3.5], we omit the details.

Lemma 4.1 shows that it is enough to find bounds on $\text{dist}(A_{n,j}, H_{n,j})$ for $j \in [n]$, where $A_{n,j}$ is the j-th column of A_n and $H_{n,j}$ is the subspace spanned by the rest of the columns. Furthermore, from the assumption on the entries of A_n it follows that one only needs to consider $j = 1$. For $j \in [n]\setminus\{1\}$ one can essentially repeat the same argument.

For a matrix A_n of i.i.d. Bernoulli entries, the first column $A_{n,1}$ is independent of $H_{n,1}$, so the desired bound on the distance essentially follows from Berry-Esséen theorem (see Lemma 3.19), upon showing that any vector in the kernel of a random matrix must be both non-dominated and incompressible. This easier case is therefore deferred to Section 5 and is dealt with during the course of the proof of Theorem 1.7. Here we will only obtain a bound on $\text{dist}(A_{n,1}, H_{n,1})$ when A_n is the adjacency matrix of either a directed or a undirected Erdős-Rényi graph.

To obtain a bound on $\text{dist}(A_{n,1}, H_{n,1})$ we derive an alternate expression for the same which is more tractable. This is done in the following extension of [45, Proposition 5.1].

Proposition 4.3 (Distance via quadratic forms)**.** *Let $\tilde{A}_n, \tilde{A}_{n,j}$ and $\tilde{H}_{n,j}$ be as in Lemma 4.1. Denote by C_n the $(n-1)\times(n-1)$ sub-matrix of $\tilde{A}_n^{\mathsf{T}}$ obtained after removing the first row and column of $\tilde{A}_n$. Furthermore, let $\boldsymbol{x}^{\mathsf{T}}, \boldsymbol{y} \in \mathbb{R}^{n-1}$ denote the first row and column of $\tilde{A}_n$ with the first entry a_{11} removed, respectively. Then we have the following:*

(i) *If C_n is non-invertible then*

$$\text{dist}(\tilde{A}_{n,1}, \tilde{H}_{n,1}) \ge \sup_{v \in \text{Ker}(C_n)\cap S^{n-1}} |\langle \boldsymbol{y}, v\rangle|,$$

where $\text{Ker}(C_n) := \{u \in \mathbb{R}^{n-1} : C_n u = 0\}$.

(ii) *If C_n is invertible then*

$$\text{dist}(\tilde{A}_{n,1}, \tilde{H}_{n,1}) = \frac{\left|\langle C_n^{-1}\boldsymbol{x}, \boldsymbol{y}\rangle - a_{11}\right|}{\sqrt{1+\left\|C_n^{-1}\boldsymbol{x}\right\|_2^2}}. \tag{4.2}$$

Proof. It follows from the definition that

$$\text{dist}(\tilde{A}_{n,1}, H_{n,1}) \ge \sup_{\boldsymbol{h}} |\langle \tilde{A}_{n,1}, \boldsymbol{h}\rangle|,$$

where the supremum is taken over all vectors $\boldsymbol{h}$ that are normal to the subspace $\tilde{H}_{n,1}$. To prove part (i) we only need to show that if $v \in \text{Ker}(C_n)$ then the vector $\binom{0}{v}$ is a vector normal to $\tilde{H}_{n,1}$. This is immediate from the definition of C_n and $\tilde{H}_{n,1}$.

When C_n is invertible and $\tilde{A}_n$ is a symmetric matrix, Proposition 4.3 was proved in [45]. A simple adaptation of this proof yields Lemma 4.3 for any square matrix $\tilde{A}_n$. We omit the details here. □

From Proposition 4.3 we see that the relevant distance has two different expressions depending on whether the $(n-1)\times(n-1)$ sub-matrix of A_n obtained upon removing the first row and column is invertible or not. In the latter case one can again use Lemma 3.19 to deduce the desired bound. Hence the treatment of that case is postponed to Section 5.

Thus the main technical result of this section is the following. For ease of writing we formulate and prove the relevant result for $(n+1)\times(n+1)$ matrices. With no loss of generality this extends to $n \times n$ matrices, possibly with slightly worse constants.

Proposition 4.4 (Distance bound)**.** *Let A_n, a matrix of size $(n+1)\times(n+1)$, be the adjacency matrix of either a directed or a undirected Erdős-Rényi graph. Let A_n be the $n \times n$ sub-matrix of A_n obtained upon deleting the first row and column of A_n. Denote $\boldsymbol{x}^{\mathsf{T}}$ and $\boldsymbol{y}$ to be the first row and column of A_n with the first common entry removed. Define*

$$\Omega_+ := \{A_n \textit{ is invertible}\},$$

and fixing $K \ge 1$ we let

$$\Omega_K^0 := \{\|A_n - \mathbb{E}A_n\| \le K\sqrt{np}\}.$$

Then there exist an absolute constant $c_{4.4}$ and another large constant $C_{4.4}$, depending only on K, such that for any $u \in \mathbb{R}$ and $\varepsilon > 0$ we have

$$\mathbb{P}\left(\left\{\frac{|\langle A_n^{-1}\boldsymbol{x}, \boldsymbol{y}\rangle - u|}{\sqrt{1+\|A_n^{-1}\boldsymbol{x}\|_2^2}} \le c_{4.4}\varepsilon\rho^2\sqrt{p}\right\} \cap \Omega_K^0 \cap \Omega_+\right) \le \varepsilon^{1/5} + \frac{C_{4.4}}{\sqrt[4]{np}}. \tag{4.3}$$

Remark 4.5. It is believed that the optimal exponent of ε in the RHS of (4.3) is one. As $\boldsymbol{x}$ and $\boldsymbol{y}$ are not independent, to obtain a bound on the probability of the event on the LHS of (4.3), we need to use a decoupling argument (see Lemma 4.7 below). Even in the case of independent $\boldsymbol{x}$ and $\boldsymbol{y}$, to apply Lemma, 3.19 one still needs to replace the denominator by some constant multiple of $\|A_n^{-1}\boldsymbol{x}\|_2$. This amounts to showing that $\|A_n^{-1}\boldsymbol{x}\|_2 \ge c$ for some $c > 0$. As the entries of A_n have a non-zero mean this poses an additional technical difficulty. These two steps together result a sub-optimal exponent of ε in the RHS of (4.3).

It is further believed that the second term in the probability bound of (4.3) can be improved to $\exp(-\bar{c}np)$ for some $\bar{c} > 0$. To improve this bound, one needs to obtain a strong estimate of the Lévy concentration function of $A_n v$ for $v \in S^{n-1}$. Such an estimate is impossible for a vector with rigid arithmetic structure. On the other hand the set of such vectors has a low metric entropy. Therefore, one needs to show that this metric entropy precisely balances the estimate on the Lévy concentration function. Putting these two pieces together, the desired better bound on the probability was obtained in [5] for sparse matrices with i.i.d. entries, for $np \ge C\log n$, for some large $C > 1$, and in [50] for symmetric sparse matrices when $p \ge n^{-c}$ for some $c \in (0,1)$. To achieve the same here for all p satisfying $np \ge \log(1/p)$ one requires new ideas. We refrain from pursuing this direction.

Remark 4.6. We point out to the reader that results analogous to Proposition 4.4 were used in [45] and [50] to control the invertibility over incompressible vectors for dense and sparse symmetric random matrices, respectively. Here to prove Proposition 4.4 we encounter additional technical difficulties to tackle the adjacency matrix of the directed Erdős-Rényi graph and also to handle the non-zero mean assumption on the entries.

Before proceeding to the proof of Proposition 4.4 let us describe the idea behind it. We note that if $\boldsymbol{x}$ and $\boldsymbol{y}$ were independent vectors with i.i.d. Bernoulli entries and if the vector $A_n^{-1}\boldsymbol{x}$ was neither dominated nor compressible then, on the event that A_n is invertible, the probability of the event

$$\left\{\frac{|\langle A_n^{-1}\boldsymbol{x}, \boldsymbol{y}\rangle - u|}{\|A_n^{-1}\boldsymbol{x}\|_2} \le \varepsilon\rho^2\sqrt{p}\right\} \tag{4.4}$$

would have been a consequence of Lemma 3.19. Therefore, applying Proposition 4.3(ii) we see that it is enough to reduce the RHS of (4.2) to an expression similar to the above. This consists of several critical steps. The first is a decoupling argument. This is done via the following lemma.

Lemma 4.7 (Decoupling)**.** *Fix any $n \times n$ matrix B_n. Suppose $\boldsymbol{z}$ and $\hat{\boldsymbol{z}}$ are random vectors of length n, with independent coordinates but not necessarily independent of each other. Further assume that for every $J \subset [n]$, $\boldsymbol{z}_J$ is independent of $\hat{\boldsymbol{z}}_{J^c}$. Let $(\boldsymbol{z}', \hat{\boldsymbol{z}}')$ be an independent copy of $(\boldsymbol{z}, \hat{\boldsymbol{z}})$. Then, for any $J \subset [n]$,*

$$\mathcal{L}\left(\langle B_n\boldsymbol{z}, \hat{\boldsymbol{z}}\rangle, \varepsilon\right)^2 \le \mathbb{P}\left(\left|\langle B_n(\boldsymbol{z}_{J^c} - \boldsymbol{z}'_{J^c}), \hat{\boldsymbol{z}}_J\rangle + \langle B_n^*(\hat{\boldsymbol{z}}_{J^c} - \hat{\boldsymbol{z}}'_{J^c}), \boldsymbol{z}_J\rangle - v\right| \le 2\varepsilon\right),$$

where v is some random vector depending on the $J^c \times J^c$ minor of B_n, and the random vectors $\boldsymbol{z}_{J^c}, \boldsymbol{z}'_{J^c}, \hat{\boldsymbol{z}}_{J^c}$, and $\hat{\boldsymbol{z}}'_{J^c}$.

Using [13, Lemma 14], in [45] (see Proposition 5.1 there), a version of Proposition 4.3 was proved when B_n a symmetric matrix and $\boldsymbol{x} = \boldsymbol{y}$. The same proof, with appropriate changes, works for a general matrix B_n and with the stated assumptions on the joint law of $\boldsymbol{x}$ and $\boldsymbol{y}$. We omit the details.

Recall that in Section 3 the invertibility over compressible and dominated vectors was proved under the general Assumption 3.1, and as seen in Remark 3.2, the assumption can be further relaxed to include skew-symmetric matrices. Since skew-symmetric matrices of odd dimension are always singular, one cannot expect to have a unified proof for all matrices satisfying this general assumption. As we will see below, the proofs for the directed and the undirected Erdős-Rényi graphs differ in choosing $J \subset [n]$ in Lemma 4.7.

So, first let us consider the case when A_n is the adjacency matrix of a directed Erdős-Rényi graph. We see that to apply Lemma 4.7 one needs to condition on A_n. Once we show that

$$1 + \|A_n^{-1}\boldsymbol{x}\|_2 \sim p^{1/2}\|A_n^{-1}\|_{\mathrm{HS}} = \Omega(1), \tag{4.5}$$

with large probability, we can replace the denominator of the RHS of (4.2) by $p^{1/2}\|A_n^{-1}\|_{\mathrm{HS}}$. This allows us to condition on A_n and then apply the decoupling lemma with an appropriate choice of the set $J \subset [n]$.

To this end, we first show that (4.5) holds when $\boldsymbol{x}$ is replaced by its centered version $\bar{\boldsymbol{x}}$. To tackle the additional difficulty of the non-zero mean we then show that the eigenvector corresponding to the largest eigenvalue of A_n is close to the vector of all ones so that $|\|A_n^{-1}\boldsymbol{x}\|_2 - \|A_n^{-1}\bar{\boldsymbol{x}}\|_2|$ is small.

Let us state the lemma showing that $\|A_n^{-1}\bar{\boldsymbol{x}}\|_2 \sim p^{1/2}\|A_n^{-1}\|_{\mathrm{HS}}$.

Lemma 4.8. *Let A_n satisfies Assumption 3.1, with p such that $np \ge \log(1/p)$. Let $\boldsymbol{x} \in \mathbb{R}^n$ be a random vector with i.i.d. $\mathrm{Ber}(p)$ entries, and $\boldsymbol{x}'$ be an independent copy of $\boldsymbol{x}$. Denote $\bar{\boldsymbol{x}} := \boldsymbol{x} - \mathbb{E}\boldsymbol{x}$. Then we have the following:*

(i) For every $\varepsilon_\star > 0$,

$$\mathbb{P}_{\boldsymbol{x}}\left(\|A_n^{-1}\bar{\boldsymbol{x}}\|_2 \le \varepsilon_\star^{-1/2} p^{1/2}(1-p)^{1/2}\|A_n^{-1}\|_{\mathrm{HS}}\right) \ge 1 - \varepsilon_\star,$$

where $\mathbb{P}_{\boldsymbol{x}}(\cdot)$ denotes the probability under the law of $\boldsymbol{x}$.

(ii) Fix $K \ge 1$. Then, for every $\varepsilon_\star > 0$,

$$\mathbb{P}\left(\left\{\|A_n^{-1}(\boldsymbol{x}-\boldsymbol{x}')\|_2 \le \varepsilon_\star p^{1/2}\rho\|A_n^{-1}\|_{\mathrm{HS}}\right\} \cap \Omega_K^0\right) \le 4C_{3.19}\left(\varepsilon_\star + \frac{Kc_{3.18}^{-1}}{\sqrt{np}}\right) + 2n^{-\bar{c}_{3.22}}, \tag{4.6}$$

where ρ as in Proposition 3.14.

The proof of Lemma 4.8 is deferred to the end of this section. The next lemma shows that A_n has a large eigenvalue and the eigenvector corresponding to that eigenvalue is close to the vector of all ones.

Lemma 4.9. *Let A_n be an (possibly random) $n \times n$ matrix and for $K \ge 1$, let*

$$\bar{\Omega}_K := \{\|A_n - p\boldsymbol{J}_n\| \le K\sqrt{np}\},$$

for some $p \in (0,1)$ such that $np \to \infty$ as $n \to \infty$, where we recall that $\boldsymbol{J}_n$ is the $n \times n$ matrix of all ones. Then on the event $\bar{\Omega}_K$, for all large n, the following hold:

(i) *There exists a real eigenvalue λ_0 of A_n such that $|\lambda_0| \ge \frac{np}{2}$.*

(ii) *We further have*

$$\left\|v_0 - \frac{1}{\sqrt{n}}\mathbf{1}\right\|_2 \le \frac{16K}{\sqrt{np}},$$

where $v_0 \in S^{n-1}$ is the eigenvector corresponding to the eigenvalue λ_0.

Equipped with Lemmas 4.8 and 4.9, one obtains (4.5), which in turn implies that one now needs to find a probability of the event

$$\left\{ \frac{|\langle A_n^{-1}(\boldsymbol{x}_{J^c} - \boldsymbol{x}'_{J^c}), \boldsymbol{y}_J\rangle + \langle (A_n^{-1})^*(\boldsymbol{y}_{J^c} - \boldsymbol{y}'_{J^c}), \boldsymbol{x}_J\rangle - v|}{\|A_n^{-1}\|_{\rm HS}} \le c\varepsilon\rho^2 p \right\}, \tag{4.7}$$

where $(\boldsymbol{x}', \boldsymbol{y}')$ is an independent copy of $(\boldsymbol{x}, \boldsymbol{y})$ and c is some small constant. As A_n is not symmetric and the first row and column, after removing the first diagonal entry, are dependent, to obtain a bound on the probability of the event in (4.7) we need a bound on the Lévy concentration function of a sum of two correlated random variables. A natural solution would be to take a $J \subset J_0 \subset [n]$ such that $\boldsymbol{x}_{J_0} \equiv 0$ so that second term in the numerator of (4.7) vanishes. Having used this trick, in order to be able to apply Lemma 3.19 we finally need to show that $v_J^\star$ has a large spread component and $\|v_J^\star\|_2$ is not too small, where

$$v^\star := \frac{A_n^{-1}(\boldsymbol{x}_{J^c} - \boldsymbol{x}'_{J^c})}{\|A_n^{-1}(\boldsymbol{x}_{J^c} - \boldsymbol{x}'_{J^c})\|_2}.$$

The existence of a $J \subset J_0 \subset [n]$ so that $v^\star$ has the desired properties is guaranteed by Lemma 3.23. Putting these pieces together one then completes the proof. Below we expand on this idea to complete the proof of Proposition 4.4 for the adjacency matrix of a directed Erdős-Rényi graph.

Proof of Proposition 4.4 for directed Erdős-Rényi graph. As mentioned above, to apply Lemma 4.7 we need to show that (4.5) holds with high probability. To this end, we begin by noting that

$$\sqrt{1 + \|A_n^{-1}\boldsymbol{x}\|_2^2} \le 2\sqrt{1 + \|A_n^{-1}\bar{\boldsymbol{x}}\|_2^2} + 2p\|A_n^{-1}\mathbf{1}\|_2,$$

where $\bar{\boldsymbol{x}} = \{\bar{x}_i\}_{i=1}^n \in \mathbb{R}^n$ and $\bar{x}_i = x_i - \mathbb{E}x_i$ for $i \in [n]$. Therefore denoting

$$\mathcal{F}_n := \left\{ p^{1/2}\|A_n^{-1}\mathbf{1}\|_2 \le \varepsilon_1^{-1/2}\|A_n^{-1}\|_{\rm HS} \right\}$$

and

$$\mathcal{G}_n := \left\{ \|A_n^{-1}\bar{\boldsymbol{x}}\|_2 \le \varepsilon_1^{-1/2} p^{1/2}(1-p)^{1/2}\|A_n^{-1}\|_{\rm HS} \right\},$$

we have that

$$\sqrt{1 + \|A_n^{-1}\boldsymbol{x}\|_2^2} \le 4\varepsilon_1^{-1/2} p^{1/2}\|A_n^{-1}\|_{\rm HS} + 2 \tag{4.8}$$

on the event $\mathcal{F}_n \cap \mathcal{G}_n$, where $\varepsilon_1 > 0$ to be determined later during the course of the proof.

We claim that $\Omega_K^0 \subset \mathcal{F}_n$. Indeed, using Lemma 4.9 we have that

$$\begin{aligned} p^{1/2}\|A_n^{-1}\mathbf{1}\|_2 &\le \sqrt{np}\cdot\|A_n^{-1}v_0\|_2 + \sqrt{np}\cdot\left\|A_n^{-1}\left(\frac{1}{\sqrt{n}}\mathbf{1} - v_0\right)\right\|_2 \\ &\le \frac{\sqrt{np}}{|\lambda_0|} + \sqrt{np}\cdot\|A_n^{-1}\|_{\rm HS}\cdot\left\|\frac{1}{\sqrt{n}}\mathbf{1} - v_0\right\|_2 \le \frac{2}{\sqrt{np}} + 32K\|A_n^{-1}\|_{\rm HS}, \end{aligned} \tag{4.9}$$

where in the last step we have used the fact that $\Omega_K^0 \subset \bar{\Omega}_{2K}$ for all large n. Using Jensen's inequality applied to the empirical measure of the square of the singular values of A_n (or equivalently using AM-HM inequality) we see that

$$\|A_n\|_{\rm HS}^2 \cdot \|A_n^{-1}\|_{\rm HS}^2 \ge n^2.$$

Since

$$\|A_n\|_{\rm HS}^2 \le 2\|A_n - \mathbb{E}A_n\|_{\rm HS}^2 + 2\|\mathbb{E}A_n\|_{\rm HS}^2 \le 2n\cdot(K^2 np) + 2(np)^2 \le 4K^2 n^2 p,$$

on the event Ω_K^0, we deduce that

$$\|A_n^{-1}\|_{\rm HS} \ge 1/(2Kp^{1/2}). \tag{4.10}$$

Plugging this bound in (4.9) and setting $\varepsilon_1 \le 10^{-4}K^{-2}$ we derive that $\Omega_K^0 \subset \mathcal{F}_n$.

Hence, estimating $\mathbb{P}(\mathcal{G}_n)$ by Lemma 4.8(i) and using (4.10) again, we obtain from (4.8) that

$$\begin{aligned}\mathbb{P}\left(\left\{\frac{|\langle A_n^{-1}\boldsymbol{x},\boldsymbol{y}\rangle - u|}{\sqrt{1+\|A_n^{-1}\boldsymbol{x}\|_2^2}} \le \varepsilon\rho^2 p^{1/2}\right\}\cap\Omega_K^0\right) \\ \le \mathbb{P}\left(\left\{\left|\langle A_n^{-1}\boldsymbol{x},\boldsymbol{y}\rangle - u\right| \le 5\varepsilon\varepsilon_1^{-1/2}p\rho^2\|A_n^{-1}\|_{\rm HS}\right\}\cap\Omega_K^0\cap\mathcal{G}_n\right)+\varepsilon_1.\end{aligned} \tag{4.11}$$

Therefore, to complete the proof it remains to find a bound on the first term in the RHS of (4.11).

Now we will apply Lemma 4.7. Recalling the fact that $\boldsymbol{x}^{\mathsf{T}}$ and $\boldsymbol{y}$ are the first row and column of A_n, after removing the first diagonal entry, using the representation (1.11) we note that

$$x_i = \theta_i \cdot \gamma_i \qquad \text{and} \qquad y_i = (1-\theta_i)\cdot\varpi_i,$$

where $\{\gamma_i\}, \{\varpi_i\}$, and $\{\theta_i\}$ are sequences of independent $\text{Ber}(2p)$, $\text{Ber}(2p)$, and $\text{Ber}(1/2)$ random variables, respectively. Set $J := \{i \in [n] : \theta_i = 0\}$. Upon conditioning on $\boldsymbol{\theta} := \{\theta_i\}_{i\in[n]}$ we see that $\boldsymbol{x}_J \equiv 0$, and $\boldsymbol{y}_J$ and $\boldsymbol{x}_{J^c}$ are distributed as i.i.d. sequence of $\text{Ber}(2p)$ random variables. Denote

$$\Upsilon_n := \mathbb{P}\left(\left\{\left|\langle A_n^{-1}\boldsymbol{x},\boldsymbol{y}\rangle - u\right| \le 5\varepsilon\varepsilon_1^{-1/2}p\rho^2\|A_n^{-1}\|_{\rm HS}\right\}\cap\Omega_K^0\right).$$

An application of Jensen's inequality and Lemma 4.7 yields that

$$\begin{aligned}\Upsilon_n^2 &\le \mathbb{E}\left[\mathcal{L}\left(\langle A_n^{-1}\boldsymbol{x},\boldsymbol{y}\rangle, 5\varepsilon\varepsilon_1^{-1/2}p\rho^2\|A_n^{-1}\|_{\rm HS}\rangle \,\Big|\, A_n,\boldsymbol{\theta}\right)^2\mathbb{I}_{\Omega_K^0}\right] \\ &\le \mathbb{E}\left[\mathbb{P}\left(\left|\langle A_n^{-1}(\boldsymbol{\gamma}-\boldsymbol{\gamma}'),\boldsymbol{\varpi}\rangle - v\right| \le 10\varepsilon\varepsilon_1^{-1/2}p\rho^2\|A_n^{-1}\|_{\rm HS}\,\Big|\, A_n,\boldsymbol{\theta}\right)\mathbb{I}_{\Omega_K^0}\right],\end{aligned} \tag{4.12}$$

where

$$\boldsymbol{\varpi} := \begin{cases}\varpi_i & i\in J\\ 0 & i\in J^c\end{cases}, \qquad \boldsymbol{\gamma} := \begin{cases}\gamma_i & i\in J^c\\ 0 & i\in J\end{cases},$$

$\boldsymbol{\gamma}'$ an independent copy of $\boldsymbol{\gamma}$, and v is some random vector depending only the $J^c \times J^c$ minor of A_n^{-1}, $\boldsymbol{\gamma}$, and $\boldsymbol{\gamma}'$.

Estimating the RHS of (4.12) relies on Lemma 3.19. To apply it, we need to bound the probability appearing there by the Lévy concentration function from this Lemma. We show that this can be done after discarding two events of a small probability.

To this end, denoting

$$\widehat{\mathcal{G}}_n := \left\{\|A_n^{-1}(\boldsymbol{\gamma}-\boldsymbol{\gamma}')\|_2 \ge \varepsilon_1 p^{1/2}\rho\|A_n^{-1}\|_{\rm HS}\right\},$$

we see that

$$\left\{\left|\langle A_n^{-1}(\boldsymbol{\gamma}-\boldsymbol{\gamma}'),\boldsymbol{\varpi}\rangle - v\right| \le 10\varepsilon\varepsilon_1^{-1/2}p\rho^2\|A_n^{-1}\|_{\rm HS}\right\}\cap\widehat{\mathcal{G}}_n \subset \left\{|\langle\boldsymbol{\xi},\boldsymbol{\varpi}\rangle - \widetilde{v}| \le 10\varepsilon\varepsilon_1^{-3/2}p^{1/2}\rho\right\}, \tag{4.13}$$

where

$$\boldsymbol{\xi} := \frac{A_n^{-1}(\boldsymbol{\gamma}-\boldsymbol{\gamma}')}{\|A_n^{-1}(\boldsymbol{\gamma}-\boldsymbol{\gamma}')\|_2} \in S^{n-1} \qquad \text{and} \qquad \widetilde{v} := \frac{v}{\|A_n^{-1}(\boldsymbol{\gamma}-\boldsymbol{\gamma}')\|_2}.$$

As $\boldsymbol{\varpi}_{J^c} \equiv 0$, to be able to apply Lemma 3.19, we have to select a set $I \subset J$ such that $\boldsymbol{\xi}_I$ has a substantial Euclidean norm and is non-dominated, with large probability.

So, we define

$$\widetilde{\mathcal{G}}_n := \left\{ \frac{\|\boldsymbol{\xi}_{[\frac{n}{4}+1:n]\cap J}\|_\infty}{\|\boldsymbol{\xi}_{[\frac{n}{4}+1:n]\cap J}\|_2} \le \frac{1}{c_{3.23}\sqrt{n}},\ \|\boldsymbol{\xi}_{[\frac{n}{4}+1:n]\cap J}\|_2 \ge \rho \right\}.$$

As the coordinates of $\boldsymbol{\varpi}_J$ are i.i.d. $\text{Ber}(2p)$, setting $I := \text{supp}(\boldsymbol{\xi}_{[\frac{n}{4}+1:n]\cap J})$, and using Lemma 3.19 we find that

$$(4.14)\quad \mathbb{P}\left(|\langle \boldsymbol{\xi}, \boldsymbol{\varpi}\rangle - \widetilde{v}| \le 10\varepsilon\varepsilon_1^{-3/2} p^{1/2}\rho \;\middle|\; A_n, \boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\gamma}' \right) \mathbb{I}_{\widetilde{\mathcal{G}}_n}$$
$$\le \mathcal{L}\left(\langle \boldsymbol{\xi}_I, \boldsymbol{\varpi}_I\rangle,\ 20\varepsilon\varepsilon_1^{-3/2} p^{1/2}(1-p)^{1/2}\|\boldsymbol{\xi}_I\|_2 \right) \mathbb{I}_{\widetilde{\mathcal{G}}_n} \le 20 C_{3.19}\left(\varepsilon\varepsilon_1^{-3/2} + \frac{1}{c_{3.23}\sqrt{np}} \right).$$

To complete the proof it remains to show that both $\widetilde{\mathcal{G}}_n$ and $\widehat{\mathcal{G}}_n$ have large probabilities. First let us show that $\mathbb{P}(\widetilde{\mathcal{G}}_n^c)$ is small. By Chernoff's bound, there exists a set Ω_γ such that on that set $|\,\text{supp}(\boldsymbol{\gamma}-\boldsymbol{\gamma}')| \le C_\star np$ and $\|\boldsymbol{\gamma}-\boldsymbol{\gamma}'\|_2 \in [1, \bar{C}np]$ for some $C_\star, \bar{C} > 0$, with $\mathbb{P}(\Omega_\gamma) \ge 1 - \exp(-\bar{c}np)$, for some constant $\bar{c} > 0$. Moreover, by Chernoff's bound again, there exists a set Ω_θ with probability at least $1 - \exp(-c_* n)$, for some $c_* > 0$, such that $\frac{3n}{8} \le |J| \le \frac{5n}{8}$ on Ω_θ. Hence, applying Lemma 3.23 we find that

$$(4.15)\quad \mathbb{P}(\widetilde{\mathcal{G}}_n^c \cap \Omega_K^0) \le \mathbb{E}\left[\mathbb{P}\left(\widetilde{\mathcal{G}}_n^c \cap \Omega_K^0 \;\middle|\; \boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\gamma}' \right) \mathbb{I}_{\Omega_\gamma \cap \Omega_\theta} \right] + \mathbb{P}(\Omega_\gamma^c) + \mathbb{P}(\Omega_\theta^c)$$
$$\le n^{-\bar{c}_{3.23}} + \exp(-\bar{c}np) + \exp(-c_* n).$$

Next, let us show that $\widehat{\mathcal{G}}_n$ has a large probability. Recall that $\boldsymbol{\gamma}$ is a random vector with independent $\text{Ber}(p)$ coordinates, and $\boldsymbol{\gamma}'$ is an independent copy of $\boldsymbol{\gamma}$. Using Lemma 4.8(ii) we obtain that

$$(4.16)\quad \mathbb{P}(\widehat{\mathcal{G}}_n^c \cap \Omega_K^0) = \mathbb{P}\left(\left\{ \|A_n^{-1}(\boldsymbol{x}-\boldsymbol{x}')\|_2 \le \varepsilon_1 p^{1/2}\rho \|A_n^{-1}\|_{\text{HS}} \right\} \cap \Omega_K^0 \right)$$
$$\le 8C_{3.19}\left(\varepsilon_1 + \frac{K c_{3.18}^{-1}}{\sqrt{np}} \right) + 2n^{-\bar{c}_{3.22}},$$

where $\boldsymbol{x}'$ is an independent copy of $\boldsymbol{x}$, establishing $\widehat{\mathcal{G}}_n$ has a large probability.

Now, combining (4.13)-(4.16), from (4.12) we derive that

$$\Upsilon_n^2 \le C(\varepsilon_1 + \varepsilon\varepsilon_1^{-3/2}) + \frac{\bar{C}}{\sqrt{np}} + n^{-c},$$

for some large constants $C, \bar{C}$, and some small constant c. This together with (4.11) now implies that

$$\mathbb{P}\left(\left\{ \frac{|\langle A_n^{-1}\boldsymbol{x}, \boldsymbol{y}\rangle - u|}{\sqrt{1+\|A_n^{-1}\boldsymbol{x}\|_2^2}} \le \varepsilon\rho^2 p^{1/2} \right\} \cap \Omega_K^0 \right) \le C(\varepsilon_1 + \varepsilon_1^{1/2} + \varepsilon^{1/2}\varepsilon_1^{-3/4}) + \frac{\bar{C}}{(np)^{1/4}} + n^{-\frac{c}{2}}.$$

Finally choosing $\varepsilon_1 = \varepsilon^{\frac{2}{5}}$ and replacing ε by ε/C^5 the proof completes. □

Next we carry out the proof for the adjacency matrix of a undirected Erdős-Rényi graph. It follows from simple modification of the same for the directed case. Hence, we only provide an outline indicating the necessary changes.

Proof of Proposition 4.4 for undirected Erdős-Rényi graph. Since in the undirected case $\boldsymbol{x} = \boldsymbol{y}$, proceeding similarly to the steps leading to (4.11) we derive that

$$\begin{aligned}(4.17)\quad \mathbb{P}\left(\left\{\frac{\left|\langle A_n^{-1}\boldsymbol{x},\boldsymbol{x}\rangle - u\right|}{\sqrt{1+\|A_n^{-1}\boldsymbol{x}\|_2^2}} \le \varepsilon\rho^2 p^{1/2}\right\} \cap \Omega_K^0\right) \\ \le \mathbb{P}\left(\left\{\left|\langle A_n^{-1}\boldsymbol{x},\boldsymbol{x}\rangle - u\right| \le 5\varepsilon\varepsilon_1^{-1/2} p\rho^2 \|A_n^{-1}\|_{\mathrm{HS}}\right\} \cap \Omega_K^0 \cap \mathcal{G}_n\right) + \varepsilon_1.\end{aligned}$$

Next set $J := \{i \in [n] : \mathfrak{d}_i = 0\}$ where $\{\mathfrak{d}_i\}_{i\in[n]}$ are i.i.d. $\mathrm{Ber}(\frac{1}{2})$. Using this choice of J we then apply Lemma 4.7 to see that

$$(4.18)\qquad \widetilde{\Upsilon}_n^2 \le \mathbb{E}\left[\mathbb{P}\left(\left|\langle A_n^{-1}(\boldsymbol{x}_{J^c} - \boldsymbol{x}'_{J^c}), \boldsymbol{x}_J\rangle - v\right| \le 5\varepsilon\varepsilon_1^{-1/2} p\rho^2 \|A_n^{-1}\|_{\mathrm{HS}} \,\Big|\, A_n, J\right) \mathbb{I}_{\Omega_K^0}\right],$$

where

$$\widetilde{\Upsilon}_n := \mathbb{P}\left(\left\{\left|\langle A_n^{-1}\boldsymbol{x},\boldsymbol{x}\rangle - u\right| \le 5\varepsilon\varepsilon_1^{-1/2} p\rho^2 \|A_n^{-1}\|_{\mathrm{HS}}\right\} \cap \Omega_K^0\right),$$

and v is some vector depending on the $J^c \times J^c$ sub-matrix of A_n, $\boldsymbol{x}_{J^c}$, and $\boldsymbol{x}'_{J^c}$. As the entries of the random vector $\boldsymbol{x}_{J^c}$ are i.i.d. $\mathrm{Ber}(\frac{p}{2})$, using Lemma 4.8(ii) we find that

$$(4.19)\qquad \mathbb{P}(\overline{\mathcal{G}}_n^c \cap \Omega_K^0) \le 16 C_{3.19}\left(\varepsilon_1 + \frac{K c_{3.18}^{-1}}{\sqrt{np}}\right) + 2n^{-\bar{c}_{3.22}},$$

where

$$\overline{\mathcal{G}}_n := \left\{\|A_n^{-1}(\boldsymbol{x}_{J^c} - \boldsymbol{x}'_{J^c})\|_2 \ge \varepsilon_1 p^{1/2}\rho\|A_n^{-1}\|_{\mathrm{HS}}\right\}.$$

As

$$\left\{\left|\langle A_n^{-1}(\boldsymbol{x}_{J^c} - \boldsymbol{x}'_{J^c}), \boldsymbol{x}_J\rangle - v\right| \le 5\varepsilon\varepsilon_1^{-1/2} p\rho^2 \|A_n^{-1}\|_{\mathrm{HS}}\right\} \cap \overline{\mathcal{G}}_n \subset \left\{\left|\langle \bar{\boldsymbol{\xi}}, \boldsymbol{x}_J\rangle - \bar{v}\right| \le 5\varepsilon\varepsilon_1^{-3/2} p^{1/2}\rho\right\},$$

where

$$\bar{\boldsymbol{\xi}} := \frac{A_n^{-1}(\boldsymbol{x}_{J^c} - \boldsymbol{x}'_{J^c})}{\|A_n^{-1}(\boldsymbol{x}_{J^c} - \boldsymbol{x}'_{J^c})\|_2} \in S^{n-1} \qquad \text{and} \qquad \bar{v} := \frac{v}{\|A_n^{-1}(\boldsymbol{x}_{J^c} - \boldsymbol{x}'_{J^c})\|_2},$$

proceeding similarly as in the proof in the directed case the remainder of this proof can be completed. We leave the details to the reader. ☐

We end this section with proofs of Lemmas 4.8 and 4.9.

Proof of Lemma 4.8. The proof of part (i) is essentially an application of Markov's inequality. To this end, we note that

$$(4.20)\qquad \|A_n^{-1}\bar{\boldsymbol{x}}\|_2^2 = \sum_{k=1}^n \langle A_n^{-1}\bar{\boldsymbol{x}}, e_k\rangle^2 = \sum_{k=1}^n \langle (A_n^{-1})^{\mathsf{T}} e_k, \bar{\boldsymbol{x}}\rangle^2 = \sum_{k=1}^n \|(A_n^{-1})^{\mathsf{T}} e_k\|_2^2 \langle w_k, \bar{\boldsymbol{x}}\rangle^2,$$

where

$$w_k := \frac{(A_n^{-1})^{\mathsf{T}} e_k}{\|(A_n^{-1})^{\mathsf{T}} e_k\|_2^2},$$

and e_k is the k-th canonical basis vector. Since $\|w_k\|_2 = 1$ and the random vector $\bar{\boldsymbol{x}}$ has zero mean with i.i.d. coordinates we have

$$\mathbb{E}_{\boldsymbol{x}}\left[\langle w_k, \bar{\boldsymbol{x}}\rangle^2\right] = \mathrm{Var}_{\boldsymbol{x}}\left(\langle w_k, \boldsymbol{x}\rangle\right) = \mathrm{Var}(x_1) = p(1-p),$$

which in turn implies that

$$\mathbb{E}_{\boldsymbol{x}}\left[\|A_n^{-1}\bar{\boldsymbol{x}}\|_2^2\right] = p(1-p)\sum_{k=1}^n \|(A_n^{-1})^{\mathsf{T}} e_k\|_2^2 = p(1-p)\|A_n^{-1}\|_{\mathrm{HS}}^2,$$

where $\mathbb{E}_{\boldsymbol{x}}$ and $\mathrm{Var}_{\boldsymbol{x}}$ denote the expectation and the variance with respect to the randomness of $\boldsymbol{x}$. The conclusion of part (i) now follows upon using Markov's inequality.

Turning to prove (ii), we denote $p_k := \|(A_n^{-1})^{\mathsf{T}} e_k\|_2^2/\|A_n^{-1}\|_{\mathrm{HS}}^2$. As $\sum_{k=1}^n p_k = 1$, proceeding as in (4.20), and applying [45, Lemma 8.3] we note that

$$\text{(4.21)}\quad \mathbb{P}\left(\|A_n^{-1}(\boldsymbol{x}-\boldsymbol{x}')\|_2 \le \varepsilon_\star p^{1/2}\rho\|A_n^{-1}\|_{\mathrm{HS}} \;\Big|\; A_n\right) = \mathbb{P}\left(\sum_{k=1}^n p_k\langle w_k, \boldsymbol{x}-\boldsymbol{x}'\rangle^2 \le \varepsilon_\star^2 p\rho^2 \;\Big|\; A_n\right)$$
$$\le 2\sum_{k=1}^n p_k \mathbb{P}\left(\langle w_k, \boldsymbol{x}-\boldsymbol{x}'\rangle^2 \le 2\varepsilon_\star^2 p\rho^2 \;\Big|v\; A_n\right).$$

As $\sum_{k=1}^n p_k = 1$, the advantage of working with the RHS of (4.21) is that it is enough to find the maximum of the probabilities under the summation. To find such a bound we would like to use Lemma 3.19. This requires to show that w_k is neither dominated nor compressible with high probability.

Turning to this task, recall that $w_k = \frac{(A_n^{-1})^{\mathsf{T}} e_k}{\|(A_n^{-1})^{\mathsf{T}} e_k\|}$. Since A_n^{T} also satisfies Assumption 3.1, applying Corollary 3.22 with $c_0^* = 1/2$, we obtain that

$$\text{(4.22)}\quad \mathbb{P}\left(\left\{\langle w_k, \boldsymbol{x}-\boldsymbol{x}'\rangle^2 \le 2\varepsilon_\star^2 p\rho^2\right\} \cap \Omega_K^0\right)$$
$$\le \mathbb{E}\left[\mathbb{P}\left(\langle w_k, \boldsymbol{x}-\boldsymbol{x}'\rangle^2 \le 2\varepsilon_\star^2 p\rho^2 \Big| A_n\right)\mathbb{I}(w_k \notin V_{1/2, c_{3.18}})\right] + n^{-\bar{c}_{3.22}}.$$

If $w_k \notin V_{1/2, c_{3.18}}$ then

$$\|(w_k)_{[n/2+1:n]}\|_2 \ge \rho \qquad \text{and} \qquad \frac{\|(w_k)_{[n/2+1:n]}\|_\infty}{\|(w_k)_{[n/2+1:n]}\|_2} \le \frac{2K}{c_{3.18}\sqrt{n}}.$$

So now we apply Lemma 3.19 to find that

$$\mathbb{P}\left(\left|\langle w_k, \boldsymbol{x}-\boldsymbol{x}'\rangle\right| \le 2\varepsilon_\star p^{1/2}\rho \Big| A_n\right)\mathbb{I}(w_k \notin V_{1/2, c_{3.18}}) \le 4C_{3.19}\left(\varepsilon_\star + \frac{K c_{3.18}^{-1}}{\sqrt{np}}\right),$$

where we have used the fact that $p \le \frac{3}{4}$. This, together with (4.22), upon taking an average over A_n, in (4.21), such that Ω_K^0 holds, yields the bound (4.6). This completes the proof of the lemma. ☐

Proof of Lemma 4.9. Denote

$$W := \left\{w \in \mathbb{R}^n : \; \|w - \mathbf{e}\|_2 \le \frac{8K}{\sqrt{np}} \quad \text{and} \quad \langle w, \mathbf{e}\rangle = 1\right\},$$

where for brevity we write $\mathbf{e} := \frac{1}{\sqrt{n}}\mathbf{1}$. Define the function $F : W \to \mathbb{R}^n$ by

$$F(x) := \frac{A_n x}{\langle A_n x, \mathbf{e}\rangle}.$$

We claim that $F(W) \subset W$. We will see below that proving this claim will imply that A_n has a large eigenvalue and the eigenvector corresponding to that large eigenvalue is close to $\mathbf{e}$.

To check the claim, note that for any $x \in \mathbb{R}^n$ we have $\langle F(x), \mathbf{e}\rangle = 1$. Therefore it remains to show that

$$\|F(x) - \mathbf{e}\|_2 \le \frac{8K}{\sqrt{np}}, \qquad \text{for all } x \in W. \tag{4.23}$$

To this end, for any $x \in W$ we write $x = \mathbf{e} + y$ where from the definition of the set W it follows that $\|y\|_2 \le \frac{8K}{\sqrt{np}}$. As $\langle x, \mathbf{e}\rangle = 1$ and $\|\mathbf{e}\|_2 = 1$ we further have that $\langle y, \mathbf{e}\rangle = 0$, which in turn implies that $A_n y = (A_n - p\boldsymbol{J}_n)y$. As

$$(A_n - p\boldsymbol{J}_n)\mathbf{e} = A_n\mathbf{e} - np\mathbf{e},$$

we deduce that

$$\|A_n\mathbf{e} - np\mathbf{e}\|_2 \le K\sqrt{np}$$

on the event $\bar{\Omega}_K$. So we obtain that

$$\|A_n x - np\mathbf{e}\|_2 \le \|A_n\mathbf{e} - np\mathbf{e}\|_2 + \|A_n - p\boldsymbol{J}_n\| \cdot \|y\|_2 \le K\sqrt{np}\left(1 + \frac{8K}{\sqrt{np}}\right) \tag{4.24}$$

on the event $\bar{\Omega}_K$. Thus using Cauchy-Schwarz inequality

$$|\langle A_n x, \mathbf{e}\rangle - np| \le K\sqrt{np}\left(1 + \frac{8K}{\sqrt{np}}\right). \tag{4.25}$$

Using the fact that $np \to \infty$ as $n \to \infty$, and the triangle inequality we also see from above that

$$|\langle A_n x, \mathbf{e}\rangle| \ge \frac{np}{2}, \tag{4.26}$$

for all large n. Combining (4.24)-(4.26), and using the triangle inequality once more, we derive that on the event $\bar{\Omega}_K$,

$$\|F(x) - \mathbf{e}\|_2 \le \frac{\|A_n x - np\mathbf{e}\|_2}{|\langle A_n x, \mathbf{e}\rangle|} + \frac{\|(\langle A_n x, \mathbf{e}\rangle - np)\mathbf{e}\|_2}{|\langle A_n x, \mathbf{e}\rangle|} \le \frac{4K}{\sqrt{np}}\left(1 + \frac{8K}{\sqrt{np}}\right) \le \frac{8K}{\sqrt{np}},$$

for all large n. This proves (4.23) and hence we have the claim that $F(W) \subset W$.

Now to show that the claim implies the existence of a real large eigenvalue we apply Brouwer fixed point theorem. It implies that there exists $w \in W$ such that

$$A_n w = \langle A_n w, \mathbf{e}\rangle w.$$

Equivalently, w is an eigenvector of A_n corresponding to the eigenvalue $\lambda_0 := \langle A_n w, \mathbf{e}\rangle$. The lower bound on $|\lambda_0|$ follows from (4.26). To complete the proof of the lemma we note that

$$|\|w\|_2 - 1| \le \|w - \mathbf{e}\|_2 \le \frac{8K}{\sqrt{np}}$$

Therefore setting $v_0 := w/\|w\|_2$ we obtain

$$\|v_0 - \mathbf{e}\|_2 \le \|w - \mathbf{e}\|_2 + \|w\|_2 \cdot \left|\frac{1}{\|w\|_2} - 1\right| \le \frac{16K}{\sqrt{np}}.$$

This finishes the proof of the lemma. □

5. Proof of Theorem 1.7

In this section we prove Theorem 1.7. First let us prove part (ii) of Theorem 1.7. We will show that the conclusion of Theorem 1.7(ii) holds under a more general set-up, namely when the entries of A_n satisfy Assumption 3.1.

Proof of Theorem 1.7(ii). The proof of (ii) is standard and is provided for a reader's convenience. We begin by noting that if A_n satisfies Assumption 3.1 it is enough to show that

$$\mathbb{P}(\Omega_{0,\mathrm{col}}^c) \le \frac{\bar{C}_{1.7}}{2\log n}, \tag{5.1}$$

where $\Omega_{0,\mathrm{col}}$ is the event that there exists zero columns in A_n.

To prove (5.1) we use Chebychev's inequality. We will show that $\mathrm{Var}(\mathscr{N}) \approx \mathbb{E}[\mathscr{N}]$, where $\mathscr{N}$ is the number of zero columns in A_n. This observation, together with the fact $\mathbb{E}[\mathscr{N}] \to \infty$ as $n \to \infty$, whenever $np \le \log(1/p)$, will show that $\mathscr{N}$ cannot deviate too much from its expectation with large probability. Then, noting that $\Omega_{0,\mathrm{col}}^c = \{\mathscr{N} = 0\}$, the desired probability bound on $\Omega_{0,\mathrm{col}}^c$ follows. Below we carry out this task.

To this end, denote $\mathbb{I}_i := \mathbb{I}_i(A_n)$ to be the indicator of the event that the i-th column of A_n is zero and therefore $\mathscr{N} = \sum_{i=1}^n \mathbb{I}_i$. It is easy to note that under Assumption 3.1 we have

$$\mathbb{E}\left[\mathscr{N}\right] = n\mathbb{P}(\mathbb{I}_1 = 1) \ge n(1-p)^n. \tag{5.2}$$

On the other hand, we see that

$$\mathrm{Var}(\mathbb{I}_i) \le \mathbb{E}\,\mathbb{I}_i \le (1-p)^{n-1}, \quad i \in [n]. \tag{5.3}$$

Using the fact that the entries of A_n satisfy Assumption 3.1 we further observe that for any $i \ne j \in [n]$ the entries of the sub-matrix of A_n with rows $([n]\backslash\{i,j\})$ and columns $\{i,j\}$ are i.i.d. $\mathrm{Ber}(p)$ random variables. Therefore

$$\begin{aligned}
\mathrm{Cov}(\mathbb{I}_i, \mathbb{I}_j) &= \mathbb{E}(\mathbb{I}_i\mathbb{I}_j) - \mathbb{E}(\mathbb{I}_i)\cdot\mathbb{E}(\mathbb{I}_j) \\
&\le \mathbb{P}\left(a_{k,\ell} = 0, (k,\ell) \in ([n]\backslash\{i,j\}) \times \{i,j\}\right) - (1-p)^{2n} \\
&= (1-p)^{2(n-2)} - (1-p)^{2n} \le Cp(1-p)^{2n},
\end{aligned} \tag{5.4}$$

for some absolute constant C, whenever $p \le 1/2$. Thus combining (5.2)-(5.4) and using Chebychev's inequality we deduce that

$$\begin{aligned}
\mathbb{P}\left(|\mathscr{N} - \mathbb{E}\mathscr{N}| \ge \frac{1}{2}\mathbb{E}\mathscr{N}\right) &\le 4\frac{n(1-p)^{n-1} + Cn^2p(1-p)^{2n}}{(\mathbb{E}\mathscr{N})^2} \\
&\le \frac{4}{(1-p)\cdot\mathbb{E}[\mathscr{N}]} + Cp \le \frac{4(1+Ce^{-1})}{(1-p)\cdot\mathbb{E}[\mathscr{N}]},
\end{aligned} \tag{5.5}$$

where the last step follows from the fact that

$$p(1-p)\mathbb{E}[\mathscr{N}] \le np(1-p)^n \le npe^{-np} \le \sup_{x\in(0,\infty)} xe^{-x} = e^{-1}.$$

To complete the argument it remains to find a suitable lower bound on $\mathbb{E}[\mathscr{N}]$. To this end, we note that the assumption $np \le \log(1/p)$ implies that $p \le 2\log n/n$. Therefore using the inequality $\log(1-x) \ge -x - x^2$ for $x \in (0, 1/2]$ we obtain that

$$\mathbb{E}[\mathscr{N}] \ge n(1-p)^n \ge np \cdot e^{-n(p+p^2)} \cdot p^{-1} \ge np \cdot e^{-\frac{4(\log n)^2}{n}} \ge np\left(1 - \frac{4(\log n)^2}{n}\right) \ge \frac{np}{2},$$

for all large n, where in the third inequality above we have again used the assumption $np \le \log(1/p)$. Thus noting that

$$\Omega_{0,\mathrm{col}}^c = \{\mathscr{N} = 0\} \subset \left\{ |\mathscr{N} - \mathbb{E}\mathscr{N}| \ge \frac{1}{2}\mathbb{E}\mathscr{N} \right\},$$

and using (5.5) we arrive at (5.1) when $p \ge \frac{\log n}{2n}$. If $p \le \frac{\log n}{2n}$, we use a different bound on $\mathbb{E}[\mathscr{N}]$:

$$\mathbb{E}[\mathscr{N}] \ge np \cdot e^{-n(p+p^2)} \cdot p^{-1} \ge np \cdot \frac{2n}{\log n} \cdot \frac{1}{2\sqrt{n}} \ge \frac{\sqrt{n}}{\log n}.$$

Proceeding as above and combining these two cases completes the proof. □

Next combining results of Sections 3-4 we finish the proof of Theorem 1.7(i).

Proof of Theorem 1.7(i). Recalling that $\Omega_K^0 = \{\|A_n - \mathbb{E}A_n\| \le K\sqrt{np}\}$, we note that for any $\vartheta > 0$,

$$\begin{aligned} &\mathbb{P}\Big(\{s_{\min}(A_n) \le \vartheta\} \cap \Omega_K^0\Big) \\ \le &\mathbb{P}\Big(\Big\{ \inf_{x\in V^c} \|A_n x\|_2 \le \vartheta \Big\} \cap \Omega_K^0\Big) + \mathbb{P}\Big(\Big\{ \inf_{x\in V} \|A_n x\|_2 \le \vartheta \Big\} \cap \Omega_K^0\Big), \end{aligned} \tag{5.6}$$

where

$$V := S^{n-1} \backslash \Big(\mathrm{Comp}(n/2, \rho) \cup \mathrm{Dom}(n/2, (c_{3.18}K^{-1})\Big), \tag{5.7}$$

and ρ as in Proposition 3.14. Since A_n, the adjacency matrix of any of the three random graph models under consideration, satisfies Assumption 3.1, using Propositions 3.14, 3.15, and 3.18, setting $y_0 = 0$, we obtain that

$$\mathbb{P}\Big(\inf_{x\in V^c} \|A_n x\|_2 \le \min\{\widetilde{c}_{3.15}, \widetilde{c}_{3.18}\}\rho\sqrt{np},\ \|A_n - \mathbb{E}A_n\| \le K\sqrt{np}\Big) \le 3n^{-\bar{c}_{3.14}}. \tag{5.8}$$

Hence, it remains to find an upper bound on the second term in the RHS of (5.6). Using Lemma 4.1, we see that to find an upper bound of

$$\mathbb{P}\Big(\Big\{ \inf_{x\in V} \|A_n x\|_2 \le c_{1.7}\varepsilon\rho^3\sqrt{\frac{p}{n}} \Big\} \cap \Omega_K^0\Big)$$

it is enough to find the same for

$$\mathbb{P}\Big(\Big\{ \mathrm{dist}(A_{n,j}, H_{n,j}) \le c_{1.7}\rho^2\sqrt{p}\varepsilon \Big\} \cap \Omega_K^0\Big) \text{ for a fixed } j,$$

where $A_{n,j}$ are now columns of A_n and $H_{n,j} := \mathrm{Span}\{A_{n,i},\ i \in [n]\backslash\{j\}\}$ (see also Remark 4.2). As A_n satisfies Assumption 3.1, it suffices consider only $j = 1$.

Turning to bound $\mathrm{dist}(A_{n,1}, H_{n,1})$, we denote C_n to be the $(n-1)\times(n-1)$ matrix obtained from A_n^{T} upon deleting its first row and column. For the adjacency matrices of directed and undirected Erdős-Rényi graphs our strategy will change depending on whether C_n is invertible or not.

Using Proposition 4.3(i) we see that

$$\begin{aligned} &\mathbb{P}\left(\left\{\mathrm{dist}(A_{n,1}, H_{n,1}) \le c_{1.7}\rho^2\sqrt{p}\varepsilon\right\} \cap \Omega_K^0 \cap \Omega_+\right) \\ &\qquad = \mathbb{P}\left(\left\{ \frac{|\langle C_n^{-1}\boldsymbol{x}, \boldsymbol{y}\rangle - a_{11}|}{\sqrt{1 + \|C_n^{-1}\boldsymbol{x}\|_2^2}} \le c_{1.7}\rho^2\sqrt{p}\varepsilon \right\} \cap \Omega_K^0 \cap \Omega_+\right), \end{aligned} \tag{5.9}$$

where by a slight abuse of notation we write $\Omega_+ := \{C_n \text{ is invertible}\}$, $\boldsymbol{x}^{\mathsf{T}}$, $\boldsymbol{y}$ are the first row and column of A_n, respectively, with the $(1,1)$-th entry a_{11} removed. As $\|C_n - \mathbb{E}C_n\| \le \|A_n - \mathbb{E}A_n\|$, using Proposition 4.4, setting $c_{1.7} \le c_{4.4}$, we see that the RHS of (5.9) is bounded by

$$\varepsilon^{1/5} + \frac{C_{4.4}}{\sqrt[4]{np}}. \tag{5.10}$$

This yields the desired bound on the event that $\operatorname{dist}(A_{n,1}, H_{n,1})$ is small on the event Ω_+. It remains to find the same on the event Ω_+^{c}. Turning to do this task, we apply Proposition 4.3(i) to obtain that

$$\mathbb{P}\Big(\Big\{\operatorname{dist}(A_{n,1}, H_{n,1}) \le c_{1.7}\rho^2\sqrt{p}\varepsilon\Big\} \cap \Omega_K^0 \cap \Omega_+^c\Big) \\ \le \mathbb{P}\Big(\Big\{\exists v \in \operatorname{Ker}(C_n) \cap S^{n-2} : \ |\langle \boldsymbol{y}, v\rangle| \le c_{1.7}\varepsilon\rho^2\sqrt{p}\Big\} \cap \Omega_K^0\Big) \tag{5.11}$$

As A_n satisfies Assumption 3.1, so does C_n. Therefore using Propositions 3.14, 3.15, and 3.18 again we obtain that

$$\mathbb{P}\left(\{\exists v \in S^{n-2} \cap \operatorname{Ker}(C_n) : v \in V^c\} \cap \Omega_K^0\right) \le 3n^{-\bar{c}_{3.14}}, \tag{5.12}$$

where we recall the definition of V from (5.7). Note that to obtain (5.12) we need to apply the propositions for a $(n-1) \times (n-1)$ matrix. This only slightly worsens the constants.

Next, by Assumption 3.1 the matrix C_n and the random vector $\boldsymbol{y}$ are independent and the coordinates of $\boldsymbol{y}$ are i.i.d. $\operatorname{Ber}(p)$. Moreover, if $v \in V$ then from the definition of V it follows that v is neither dominated nor compressible. Hence, conditioning on such a realization of $v \in S^{n-2} \cap \operatorname{Ker}(C_n)$, applying Lemma 3.19, and finally taking an average over such choices of v we obtain

$$\mathbb{P}\Big(\Big\{\exists v \in \operatorname{Ker}(C_n) \cap V : \ |\langle \boldsymbol{y}, v\rangle| \le c_{1.7}\varepsilon\rho^2\sqrt{p}\Big\} \cap \Omega_K^0\Big) \le C_{3.19}\left(\varepsilon + \frac{2K}{c_{3.18}\sqrt{np}}\right). \tag{5.13}$$

To complete the proof for the adjacency matrices of the directed and undirected Erdős-Rényi graphs we simply take $c_0 = \frac{1}{2}$ and $C_0 = 1$ in Theorem 1.14 and then set $K = C_{1.14}$. Now combining (5.8)-(5.13), applying Theorem 1.14, and substituting the bounds in (5.6) we arrive at (1.13) when $A_n = \operatorname{Adj}(\mathsf{G}(n, p_n))$ or $\operatorname{Adj}(\overrightarrow{\mathsf{G}}(n, p_n))$.

Next, we need to obtain a lower bound on $s_{\min}$ for the adjacency matrix of random bipartite random graphs. We recall that it is enough to establish the same for a matrix with i.i.d. $\operatorname{Ber}(p)$ entries. For the latter matrix A_n we note that $\operatorname{dist}(A_{n,1}, H_{n,1}) \ge |\langle A_{n,1}, v\rangle|$ for any $v \in \operatorname{Ker}(\widetilde{A}_n) \cap S^{n-1}$ where $\widetilde{A}_n$ is the $(n-1)\times n$ matrix whose rows are the columns $A_{n,2}, \ldots, A_{n,n}$. Since the entries of A_n are independent, we apply Propositions 3.14, 3.15, and 3.18 for the matrix $\widetilde{A}_n$ (although these were proved for square matrices, they have a simple extension for $\widetilde{A}_n$; see also Remark 3.21) to conclude that any $v \in \operatorname{Ker}(\widetilde{A}_n) \cap S^{n-1}$ must also be in V with high probability. Then arguing similarly as in (5.13), followed by an application of Theorem 1.14 to bound the probability of $(\Omega_K^0)^c$, we arrive at (1.13). This completes the proof of the theorem. □

6. Bound on the spectral norm

In this short section we prove Theorem 1.14 which yields the desired bound on $\|A_n - \mathbb{E}A_n\|$.

Proof of Theorem 1.14. The proof consists of two parts. We will show that $\|A_n - \mathbb{E}A_n\|$ concentrates near its mean and then find bounds on $\mathbb{E}\|A_n - \mathbb{E}A_n\|$. First let us derive the concentration of $\|A_n - \mathbb{E}A_n\|$. Since $a_{i,j}$ may depend on $a_{j,i}$ we split A_n into its upper and lower triangular part (excluding the diagonal), denoted hereafter by A_n^U and A_n^L, respectively, and work with them separately.

The function $\|A_n^U - \mathbb{E}A_n^U\|$ when viewed as a function from $\mathbb{R}^{n(n-1)/2}$ to $\mathbb{R}$ is a 1-Lipschitz, quasi-convex function. So using Talagrand's inequality (see [9, Theorem 7.12]) we obtain that for any $t > 0$,

$$\mathbb{P}\left(\left|\|A_n^U - \mathbb{E}A_n^U\| - \mathbb{M}_n\right| \ge t\right) \le 4\exp(-t^2/4), \tag{6.1}$$

where $\mathbb{M}_n$ is the median of $\|A_n^U - \mathbb{E}A_n^U\|$. Using integration by parts from (6.1) it also follows that $|\mathbb{E}\|A_n^U - \mathbb{E}A_n^U\| - \mathbb{M}_n| \le C^*$ for some absolute constant C^*. Since A_n satisfies Assumption 3.1, so does A_n^{T}. Hence, proceeding similarly as above, we find that same holds for A_n^L. As the entries of A_n are $\{0,1\}$-valued it follows that $\|A_n^D - \mathbb{E}A_n^D\| \le 1$, where A_n^D is the diagonal part of A_n. Hence, using the triangle inequality we deduce that

$$\mathbb{P}\left(\|A_n - \mathbb{E}A_n\| \le \mathbb{E}\|A_n^U - \mathbb{E}A_n^U\| + \mathbb{E}\|A_n^L - \mathbb{E}A_n^L\| + \widetilde{C}\sqrt{np}\right) \le \exp(-C_0 \log n), \tag{6.2}$$

for some large constant $\widetilde{C}$.

Now it remains to show that $\mathbb{E}\|A_n^\dagger - \mathbb{E}A_n^\dagger\| \le \bar{C}\sqrt{np}$ for $\dagger \in \{U, L\}$. To this end, let A_n' be an independent copy of A_n and R_n be a $n \times n$ symmetric matrix consisting of independent Rademacher random variables. Since, the entries of $A_n - A_n'$ have a symmetric distribution, applying Jensen's inequality we obtain that,

$$\mathbb{E}\|A_n^U - \mathbb{E}A_n^U\| \le \mathbb{E}\|A_n^U - A_n'^U\| = \mathbb{E}\|D_n^U \odot R_n\|, \tag{6.3}$$

where we denote $D_n := A_n - A_n'$, and $D_n \odot R_n$ denotes the Hadamard product of D_n and R_n. Next, let us denote G_n to be a $n \times n$ symmetric matrix with independent standard Gaussian random variables and $|G_n|$ to be the matrix constructed from G_n by taking absolute value of each of its entries. We write $\mathbb{E}_r$ and $\mathbb{E}_g$ to denote the expectations with respect to R_n and G_n respectively. Therefore, applying Jensen's inequality again

$$\mathbb{E}_r\|D_n^U \odot R_n\| = \sqrt{\frac{\pi}{2}}\mathbb{E}_r\|D_n^U \odot R_n \odot \mathbb{E}|G_n|\| \le \sqrt{\frac{\pi}{2}}\mathbb{E}_r\mathbb{E}_g\|D_n^U \odot R_n \odot |G_n|\| = \sqrt{\frac{\pi}{2}}\mathbb{E}_g\|D_n^U \odot G_n\|. \tag{6.4}$$

This implies that it is enough to bound $\mathbb{E}\|D_n^U \odot G_n\|$. Using [2, Theorem 1.1] we obtain that

$$\mathbb{E}_g\|D_n^U \odot G_n\| \le C\left[\sigma + \sqrt{\log n}\right], \tag{6.5}$$

where

$$\sigma := \max\left\{\max_i \sqrt{\sum_j \mathfrak{d}_{i,j}^2}, \max_j \sqrt{\sum_i \mathfrak{d}_{i,j}^2}\right\},$$

$\mathfrak{d}_{i,j}$ is the (i,j)-th entry of D_n, and C is an absolute constant. Using Chernoff bound and the union bound we note that there exists a constant C' depending only c_0 (recall $p \ge c_0 \frac{\log n}{n}$), such that

$$\mathbb{P}(\Omega) \ge 1 - n^{-2}, \quad \text{where} \quad \Omega := \{\sigma \le C'\sqrt{np}\}.$$

Therefore fixing a realization of D_n such that $\sigma \in \Omega$ from (6.4)-(6.5) we find

$$\mathbb{E}_r\|D_n^U \odot R_n\| \le \frac{\bar{C}}{2}\sqrt{np},$$

for some constant $\bar{C}$, depending only on c_0. On the other hand noting that the entries of $D_n \odot R_n$ are $\{-1,0,1\}$ valued it is easily follows that $\|D_n \odot R_n\| \le n$. So

$$\mathbb{E}\|D_n^U \odot R_n\| \le \mathbb{E}\left[\mathbb{I}(\Omega)\mathbb{E}_r\|D_n^U \odot R_n\|\right] + n\mathbb{P}(\Omega^c) \le \bar{C}\sqrt{np}.$$

Hence, from (6.3) we now have

$$\mathbb{E}\|A_n^U - \mathbb{E}A_n^U\| \le \bar{C}\sqrt{np}.$$

Same bound holds for $\mathbb{E}\|A_n^L - \mathbb{E}A_n^L\|$. Therefore the proof now finishes from (6.2). □

Appendix A. Structural properties of the adjacency matrices of sparse graphs

In this section we prove that certain structural properties of A_n, as listed in Lemma 3.7, hold with high probability when A_n satisfies Assumption 3.1 with p such that $np \ge \log(1/\bar{C}p)$, for some $\bar{C} \ge 1$. We also show that under the same assumption we have bounds on the number of light columns of A_n, namely we prove Lemma 3.13.

First let us provide the proof of Lemma 3.13.

Proof of Lemma 3.13. The proof is a simple application of Chernoff bound and Markov's inequality.

Since the entries of A_n satisfies Assumption 3.1, using Stirling's approximation we note that

$$\mathbb{P}(\text{col}_j(A_n) \text{ is light}) \le \sum_{\ell=0}^{\delta_0 np} \binom{n-1}{\ell} p^\ell (1-p)^{n-1-\ell} \le 2\delta_0 np \left(\frac{e}{\delta_0}\right)^{\delta_0 np} \cdot \exp(-p(n-\delta_0 np))$$

$$\le \exp\left(-np\left[1 - \delta_0 p - \delta_0 \log\left(\frac{2e}{\delta_0}\right)\right]\right), \tag{A.1}$$

where in the second inequality we have used the fact that $p \le 1/4$. Therefore, for $np \ge C\log n$, with C large, using the union bound we find $\mathbb{E}[|\mathcal{L}(A_n)|] < 1/n$. Hence by Markov's inequality we deduce that

$$\mathbb{P}(\mathcal{L}(A_n) \ne \varnothing) = \mathbb{P}(|\mathcal{L}(A_n)| \ge 1) \le \mathbb{E}[|\mathcal{L}(A_n)|] \le 1/n.$$

To prove the upper bound on the cardinality of $\mathcal{L}(A_n)$ we note that the assumption $np \ge \log(1/\bar{C}p)$ implies that $np \ge (1-\delta)\log n$, for any $\delta > 0$, for all large n. Therefore, using (A.1) and Markov's inequality, setting $\delta = \frac{1}{9}$, we find that for $np \le 2\log n$,

$$\mathbb{P}(|\mathcal{L}(A_n)| \ge n^{\frac{1}{3}}) \le n^{-\frac{1}{3}}\mathbb{E}|\mathcal{L}(A_n)| \le n^{\frac{2}{3}} \cdot n^{-\frac{8}{9}} \cdot n^{2\delta_0 p + 2\delta_0 \log\left(\frac{2e}{\delta_0}\right)} \le n^{-\frac{1}{9}},$$

for all large n, whenever δ_0 is chosen sufficiently small. For p such that $2\log n \le np \le C_{3.13}\log n$ we note from (A.1) that

$$\mathbb{P}(\text{col}_j(A_n) \text{ is light}) \le \frac{1}{n}, \qquad j \in [n].$$

Therefore, an union bound followed by Markov's inequality yield the desired result. □

Proof of Lemma 3.7. We will show that each of the six properties of the event $\Omega_{3.7}$ hold with probability at least $1 - Cn^{-2\bar{c}_{3.7}}$, for some constant $C > 0$. Then, taking a union bound the desired conclusion would follow.

First let us start with the proof of (1). Since the inequality $np \ge \log(1/\bar{C}p)$ implies that $np \ge \log n/2$, for all large n, it follows from Chernoff bound that property (1) of the event $\Omega_{3.7}$ holds with probability at least $1 - 1/n$, for all large n. We omit the details.

Next let us prove that property (2) of $\Omega_{3.7}$ holds with high probability. For $(i,j) \in \binom{[n]}{2}$ and $k \in [n]$ denote by $\Omega_{(i,j),k}$ the event that the columns $\text{col}_i(A_n), \text{col}_j(A_n)$ are light and $a_{k,i}, a_{k,j} \ne 0$. Note that the event that two light columns intersect is contained in the event $\cup_{i,j,k}\Omega_{(i,j),k}$. Therefore, we need to find bounds $\mathbb{P}(\Omega_{(i,j),k})$. Since the entry $a_{i,j}$ may depend on $a_{j,i}$ we need to consider the cases $k \in [n]\backslash\{i,j\}$ and $k \in \{i,j\}$ separately.

First let us fix $k \in [n]\backslash\{i,j\}$. We note that

$$\Omega_{(i,j),k} \subset \{a_{k,i} = a_{k,j} = 1, |\operatorname{supp}(\text{col}_i(A_n))\backslash\{i,j\}|, |\operatorname{supp}(\text{col}_j(A_n))\backslash\{i,j\}| \le \delta_0 np\}.$$

Therefore, recalling that under Assumption 3.1 the entries of the sub-matrix of A_n indexed by $([n]\backslash\{i,j\}) \times \{i,j\}$ are i.i.d. Ber(p) we obtain that

$$\mathbb{P}(\Omega_{(i,j),k}) \le p^2 \exp\left(-2np\left[1-\delta_0 p - \delta_0 \log\left(\frac{2e}{\delta_0}\right)\right]\right) =: q,$$

for all large n, where we have proceeded similarly as in (A.1) to bound the probability of the event

$$\{|\operatorname{supp}(\operatorname{col}_i(A_n))\backslash\{i,j\}|, |\operatorname{supp}(\operatorname{col}_j(A_n))\backslash\{i,j\}| \le \delta_0 np\}.$$

Since $np \ge \log(1/\bar{C}p)$ an application of the union bound shows that

$$\mathbb{P}\left(\bigcup_{i\neq j\in[n], k\notin\{i,j\}} \Omega_{(i,j),k}\right) \le n\cdot\binom{n}{2} q \le \frac{p^{-1}}{2}e^{-np}\cdot(np)^3\cdot\exp\left(-np\left[1-2\delta_0 p - 2\delta_0\log\left(\frac{2e}{\delta_0}\right)\right]\right)$$

$$\le \frac{\bar{C}}{2}\cdot(np)^3\cdot\exp\left(-np\left[1-2\delta_0 p-2\delta_0\log\left(\frac{2e}{\delta_0}\right)\right]\right) \le n^{-c}, \tag{A.2}$$

for some absolute constant c and all large n, where we use that $np \ge \log n/2$, which as already seen is a consequence of the assumption $np \ge \log(1/\bar{C}p)$.

Next let us consider the case $k \in \{i,j\}$. Without loss of generality, let us assume that $k = i$. We see that

$$\Omega_{(i,j),i} \subset \{a_{i,j}=1, |\operatorname{supp}(\operatorname{col}_i(A_n))\backslash\{i,j\}|, |\operatorname{supp}(\operatorname{col}_j(A_n))\backslash\{i,j\}| \le \delta_0 np\}.$$

Hence proceeding same as above we deduce

$$\mathbb{P}\left(\bigcup_{i\neq j\in[n], k\in\{i,j\}} \Omega_{(i,j),k}\right) \le 2\cdot\binom{n}{2}\cdot p\cdot\exp\left(-2np\left[1-\delta_0 p-\delta_0\log\left(\frac{2e}{\delta_0}\right)\right]\right)$$

$$\le p^{-1}e^{-np}\cdot(np)^2\cdot\exp\left(-np\left[1-2\delta_0 p-2\delta_0\log\left(\frac{2e}{\delta_0}\right)\right]\right) \le n^{-c}. \tag{A.3}$$

So combining the bounds of (A.2)-(A.3) we conclude that property (2) of $\Omega_{3.7}$ holds with probability at least $1 - n^{-2\bar{c}_{3.7}}$.

Now let us prove that (3) holds with high probability. We let $j \in [n]$, $I = (i_1, \ldots, i_{r_0}) \in \binom{[n]\backslash\{j\}}{r_0}$, and $k_1, \ldots, k_{r_0} \in [n]$, for some absolute constant r_0 to be determined during the course of the proof. Denote by $\Omega_{j,I,(k_1,\ldots,k_{r_0})}$ the event that all the columns indexed by I are light, and for any $i_\ell \in I$, $k_\ell \in \operatorname{supp}(\operatorname{col}_{i_\ell}(A_n)) \cap \operatorname{supp}(\operatorname{col}_j(A_n))$. Equipped with this notation we see that the event that there exists a column such that its support intersects with the supports of at least r_0 light columns is contained in the event $\cup_{j;I;k_\ell,\ell\in[r_0]}\Omega_{j,I,(k_1,k_2,\ldots,k_{r_0})}$.

Since all the columns indexed by I are light, applying property (2) it follows that $\{k_\ell\}_{\ell=1}^{r_0}$ are distinct. Therefore, for matrices with independent entries (3) follows upon bounding the probability of the events

$$\left|\operatorname{supp}(\operatorname{col}_{i_\ell}(A_n))\backslash\{k'_\ell\}_{\ell'=1}^{r_0}\right| \le \delta_0 np, \quad \ell\in[r_0]$$

and

$$a_{k_\ell,j} = a_{k_\ell,i_\ell} = 1, \qquad \ell \in [r_0],$$

followed a union bound. Recall that under Assumption 3.1 the entry $a_{i,j}$ may only depend on $a_{j,i}$ for $i,j \in [n]$. Therefore, to carry out this scheme for matrices satisfying Assumption 3.1 we additionally need to show that the support of $\operatorname{col}_j(A_n)$ is almost disjoint from the set of light columns with high probability, so that we can omit the relevant diagonal block to extract a sub-matrix with jointly independent entries.

To this end, we claim that

$$\mathbb{P}\left(\exists j \in [n] : |\operatorname{supp}(\operatorname{col}_j(A_n)) \cap \mathcal{L}(A_n)| \geq 3\right) \leq n^{-c'}, \tag{A.4}$$

for some $c' > 0$. To establish (A.4) we fix $j \in [n]$ and note that

$$\begin{aligned}
&\{|\operatorname{supp}(\operatorname{col}_j(A_n)) \cap \mathcal{L}(A_n)| \geq 3, |\operatorname{supp}(\operatorname{col}_j(A_n))| \leq C_{3.7} np\} \\
\subset &\big\{\exists k \text{ with } 2 \leq k \leq C_{3.7} np, \text{ and } i_1, i_2, \ldots, i_k \in [n] \backslash \{j\} \text{ distinct such that} \\
&\qquad |\operatorname{supp}(\operatorname{col}_{i_\ell}(A_n)) \backslash \{i_1, i_2 \ldots, i_k, j\}| \leq \delta_0 np, \ell = 1, 2, \ldots, k\big\}.
\end{aligned}$$

For ease of writing, let us denote

$$q' := \exp\left(-np\left[1 - \delta_0 p - C_{3.7} p - \delta_0 \log\left(\frac{2e}{\delta_0}\right)\right]\right).$$

By Assumption 3.1 the entries $\{a_{i',j'}\}$ for $(i', j') \in \{i_\ell\}_{\ell=1}^k \times ([n] \backslash (\{i_\ell\}_{\ell=1}^k \cup \{j\})$ are jointly independent $\text{Ber}(p)$ random variables. Therefore applying Stirling's approximation once more, and proceeding similarly as in (A.1) we find that

$$\begin{aligned}
&\mathbb{P}\Big(|\operatorname{supp}(\operatorname{col}_j(A_n)) \cap \mathcal{L}(A_n)| \geq 3, |\operatorname{supp}(\operatorname{col}_j(A_n))| \leq C_{3.7} np \Big| \operatorname{col}_j(A_n)\Big) \\
&\leq \sum_{k=2}^{C_{3.7} np} \binom{C_{3.7} np}{k} q'^k \leq \sum_{k \geq 2} \left(\frac{e C_{3.7} np}{k}\right)^k \cdot q'^k \\
&\qquad \leq e^{-np} p^{-1} \cdot p \cdot (e C_{3.7} np)^2 \cdot \exp\left(-np\left[1 - 2\delta_0 p - 2C_{3.7} p - 2\delta_0 \log\left(\frac{2e}{\delta_0}\right)\right]\right).
\end{aligned}$$

Since by Lemma 3.13 we see that $\mathcal{L}(A_n) = \varnothing$ with high probability when $p \geq \frac{C_{3.13} \log n}{n}$. Without loss of generality, we therefore assume that $p \leq \frac{C_{3.13} \log n}{n}$. So, by the union bound over j, using the fact that $np \geq \log(1/\bar{C} p)$ and property (1) of the event $\Omega_{3.7}$ we have that, for all large n,

$$\begin{aligned}
\mathbb{P}\left(\exists j \in [n] : |\operatorname{supp}(\operatorname{col}_j(A_n)) \cap \mathcal{L}(A_n)| \geq 3\right) &\leq \bar{C} \cdot (e C_{3.7} np)^3 \cdot \exp\left(-np(1 - \delta)\right) + 1/n \\
&\leq 2 \exp(-np(1 - 2\delta)),
\end{aligned}$$

for some $\delta > 0$. This establishes the claim (A.4).

Equipped with (A.4) we turn to proving (3). Using (A.4) we see that excluding a set of probability at most $n^{-c'}$, for any $j, I, (k_1, \ldots, k_{r_0})$ such that $\Omega_{j,I,(k_1,\ldots,k_{r_0})}$ occurs, we can find $\ell_1, \ldots, \ell_{r_0 - 3}$ with $k_{\ell_s} \in [n] \backslash (\mathcal{L}(A_n) \cup \{j\}) \subset [n] \backslash (I \cup \{j\})$ for all $s = 1, 2, \ldots, r_0 - 3$. For such k_{ℓ_s}, all events $|\operatorname{supp}(\operatorname{col}_{i_{\ell_s}}(A_n)) \backslash (I \cup \{j\})| \leq \delta_0 np$ and $a_{k_{\ell_s}, j} = a_{k_{\ell_s}, i_{\ell_s}} = 1$ with $s = 1, 2, \ldots, r_0 - 3$ are independent. Denote for brevity

$$\bar{q} := \exp\left(-np\left[1 - \delta_0 p - \delta_0 \log\left(\frac{2e}{\delta_0}\right)\right]\right).$$

Note that under the assumption $np \geq \log(1/\bar{C} p)$ we have $\bar{q} \leq \exp(-\log n/2)$ for all large n. Hence, recalling Assumption 3.1, using (A.4) and property (2) of $\Omega_{3.7}$, and proceeding similarly as in (A.1) once again we see that

$$\mathbb{P}\left(\bigcup_{\substack{j,k_1,,\ldots,k_{r_0}\in[n]\\ I\in\binom{[n]\setminus\{j\}}{r_0}}}\Omega_{j,I,(k_1,,\ldots,k_{r_0})}\right) \tag{A.5}$$
$$\le \sum_{\substack{j,k_1,,\ldots,k_{r_0}\in[n]\\ I\in\binom{[n]\setminus\{j\}}{r_0}}}\prod_{s=1}^{r_0-3}\mathbb{P}\left(|\operatorname{supp}(\operatorname{col}_{i_{\ell_s}}(A_n))\setminus(I\cup\{j\})|\le\delta_0 np\right)\cdot\mathbb{P}\left(a_{k_{\ell_s},j}=a_{k_{\ell_s},i_{\ell_s}}=1\right)+n^{-c_0}$$
$$\le n^{r_0+1}\binom{n-1}{r_0}p^{2(r_0-3)}\cdot\bar{q}^{r_0-3}+n^{-c'}\le(np)^{2(r_0-3)}\cdot n^7\cdot\bar{q}^{r_0-3}+n^{-c'}\le n^{-\bar{c}}+n^{-c_0},$$

for some $\bar{c}, c_0 > 0$, where the last step follows upon choosing r_0 such that $r_0 - 3 > 15$. This completes the proof of property (3).

Next let us show that (4) holds with high probability. First we will prove that for any $j \in [n]$ such that $\operatorname{col}_j(A_n)$ is normal we have

$$\left|\operatorname{supp}(\operatorname{col}_j(A_n))\cap\left(\bigcup_{i\in\mathcal{L}(A_n)}\operatorname{supp}(\operatorname{col}_i(A_n))\right)\right|\le\frac{\delta_0}{64}np, \tag{A.6}$$

with high probability. Note that the difference between (A.6) and property (4) of $\Omega_{3.7}$ is that in (A.6) it is claimed that for any $j \in [n]$ such that $\operatorname{col}_j(A_n)$ is normal its support does not have a large intersection with that of light columns. To establish property (4) we need to strengthen the above to deduce that one can replace the matrix A_n by its folded version on the LHS of (A.6) with the loss of factor of four in its RHS.

Turning to prove (A.6), we see that if (3) holds then given any $j \in [n]$ there exists only r_0 light columns $\operatorname{col}_{i_1}(A_n),,\ldots,,\operatorname{col}_{i_{r_0}}(A_n)$ such that their supports intersect that of $\operatorname{col}_j(A_n)$. Hence,

$$\left\{\exists j\in[n]\setminus\mathcal{L}_n(A):\left|\operatorname{supp}(\operatorname{col}_j(A_n))\cap\left(\bigcup_{i\in\mathcal{L}(A_n)}\operatorname{supp}(\operatorname{col}_i(A_n))\right)\right|\ge\frac{\delta_0}{64}np\right\}\cap\{(3)\text{ holds}\}$$
$$\subset\left\{\exists i\ne j\in[n]:|\operatorname{supp}(\operatorname{col}_j(A_n))\cap\operatorname{supp}(\operatorname{col}_i(A_n))|\ge\frac{\delta_0}{64r_0}np\right\}. \tag{A.7}$$

Since by (1) we have that $|\operatorname{supp}(\operatorname{col}_j(A_n))| \le C_{3.7}np$, using Stirling's approximation and a union bound we show that the event on the RHS of (A.7) holds with small probability.

Indeed, for $i \ne j \in [n]$, denoting

$$\bar{\Omega}_{i,j} := \left\{|\operatorname{supp}(\operatorname{col}_j(A_n))\cap\operatorname{supp}(\operatorname{col}_i(A_n))|\ge\frac{\delta_0}{64r_0}np\right\},$$

and using the fact that property (1) holds with high probability we deduce that

$$\mathbb{P}\left(\bigcup_{i\neq j\in[n]}\bar{\Omega}_{i,j}\right)\le\sum_{i\neq j\in[n]}\mathbb{E}\left[\mathbb{P}\left(\bar{\Omega}_{i,j}\cap\{|\operatorname{supp}(\operatorname{col}_j(A_n))|\le C_{3.7}np\}\,\Big|\,\operatorname{col}_j(A_n)\right)\right]+n^{-1}$$

$$\le\binom{n}{2}\cdot\binom{C_{3.7}np}{\frac{\delta_0}{64r_0}np}p^{\frac{\delta_0}{64r_0}np}+n^{-1}\le n^2\cdot\left(\frac{eC_{3.7}64r_0p}{\delta_0}\right)^{\frac{\delta_0}{64r_0}np}+n^{-1}\le 2n^{-1}, \tag{A.8}$$

for all large n. Thus combining (A.7)-(A.8) and applying property (1) of the event $\Omega_{3.7}$ we establish that (A.6) holds with probability at least $1-n^{-\widetilde{c}}$ for some $\widetilde{c}>0$.

As mentioned above, to show that property (4) holds with high probability we need to strengthen (A.6). To this end, recalling the definition of the folded matrix (see Definition 3.5) we note that $k\in\operatorname{supp}(\operatorname{col}_i(\operatorname{fold}(A_n))\cap\operatorname{supp}(\operatorname{col}_j(\operatorname{fold}(A_n))$ implies that

$$k\in\operatorname{supp}_{\mathfrak{u}}(\operatorname{col}_i(A_n))\cap\operatorname{supp}_{\mathfrak{v}}(\operatorname{col}_j(A_n))$$

for some $\mathfrak{u},\mathfrak{v}\in\{1,2\}$, where for any $\ell\in[n]$.

$$\operatorname{supp}_1(\operatorname{col}_\ell(A_n)):=\operatorname{supp}(\operatorname{col}_\ell(A_n))\cap[\mathfrak{n}],$$

$$\operatorname{supp}_2(\operatorname{col}_\ell(A_n)):=(\operatorname{supp}(\operatorname{col}_\ell(A_n))\cap[\mathfrak{n}+1,2\mathfrak{n}])-\mathfrak{n},$$

and for any set $S\subset[n]$ and $k\in\mathbb{Z}$ we denote $S+k:=\{x+k:x\in S\}$. Using the observation we see that it suffices to show that

$$\left|\operatorname{supp}_{\mathfrak{u}}(\operatorname{col}_j(A_n))\cap\left(\bigcup_{i\in\mathcal{L}(A_n)}\operatorname{supp}_{\mathfrak{v}}(\operatorname{col}_i(A_n))\right)\right|\le\frac{\delta_0}{64}np, \tag{A.9}$$

with high probability, for all $\mathfrak{u},\mathfrak{v}\in\{1,2\}$. If $\mathfrak{u}=\mathfrak{v}$ then (A.9) is an immediate consequence of (A.6). It remains to prove (A.9) for $\mathfrak{u}\neq\mathfrak{v}$. Let us consider the case $\mathfrak{u}=1$ and $\mathfrak{v}=2$. From (A.4) we have

$$\mathbb{P}\left(\exists j\in[n]:|\operatorname{supp}_1(\operatorname{col}_j(A_n))\cap\mathcal{L}(A_n)|\ge 3\right)\le n^{-c'}.$$

Therefore, proceeding similarly as in the steps leading to (A.5) we deduce that, with the desired high probability, for any $j\in[n]$, such that $\operatorname{col}_j(A_n)$ is a normal column, there are at most r_0 light columns $\{\operatorname{col}_{i_\ell}(A_n)\}_{\ell=1}^{r_0}$ so that $\operatorname{supp}_1(\operatorname{col}_j(A_n))\cap\operatorname{supp}_2(\operatorname{col}_{i_\ell}(A_n))\neq\varnothing$. Now arguing similarly as in the proof of (A.6) we derive (A.9) for $\mathfrak{u}=1$ and $\mathfrak{v}=2$. The proof of the other case is similar and hence is omitted.

Next we show that (5) holds with high probability. We first fix an $I\subset[n]$ with $2\le|I|\le c_{3.7}p^{-1}$ and derive that (5) holds with certain probability for each such choice of I and then take an union over I.

Since the entry $a_{i,j}$ may depend on $a_{j,i}$, for $i\neq j$, to derive that (5) holds with the desired probability we need to split it into two cases. Namely, the off-diagonal and the diagonal blocks require separate arguments. First we consider the off-diagonal block.

To this end, define the random variables

$$\eta_i:=\max(|\{j\in I:\ \mathfrak{a}_{i,j}\neq 0\}|-1,0),\qquad i\in[\mathfrak{n}]\backslash\bar{I},$$

where we recall $\bar{I}:=\bar{I}(I):=\{j\in[\mathfrak{n}]:j\in I\text{ or }j+\mathfrak{n}\in I\}\subset[\mathfrak{n}]$, $\mathfrak{n}:=\lfloor n/2\rfloor$, and $\mathfrak{a}_{i,j}$ denotes the (i,j)-th entry of $\operatorname{fold}(A_n)$. Observe that

$$\left|\bigcup_{j\in I}\left(\operatorname{supp}(\operatorname{col}_j(\operatorname{fold}(A_n)))\backslash\bar{I}\right)\right|=\sum_{j\in I}\left|\operatorname{supp}(\operatorname{col}_j(\operatorname{fold}(A_n)))\backslash\bar{I}\right|-\sum_{i\in[\mathfrak{n}]\backslash\bar{I}}\eta_i.$$

To prove (5) we need to show that $\sum \eta_i$ cannot be too large with large probability. To show the latter we use the standard Laplace transform method.

Note that

$$\mathfrak{a}_{i,j} = \xi_{i,j} \cdot \delta_{i,j}, \quad i \in [\mathfrak{n}] \backslash \bar{I}, j \in I,$$

where $\{\xi_{i,j}\}$ are i.i.d. Rademacher random variables, $\delta_{i,j}$ are i.i.d. $\text{Ber}(\mathfrak{p})$ random variables, and $\mathfrak{p} := 2p(1-p)$. Therefore,

$$\mathbb{P}\{\eta_i = \ell\} \le \binom{|I|}{\ell+1} \mathfrak{p}^{\ell+1}, \quad \ell \in \mathbb{N}.$$

Thus, for any $\lambda > 0$ such that $e^\lambda \mathfrak{p}|I| \le 1$, we have

$$\mathbb{E}\left(e^{\lambda \eta_i}\right) \le 1 + \sum_{\ell=1}^{\infty} \left(e^\lambda\right)^\ell \mathfrak{p}^{\ell+1} |I|^{\ell+1} ((\ell+1)!)^{-1} \le 1 + e\mathfrak{p}|I|,$$

and hence

$$\mathbb{P}\left\{\sum_{i \in [\mathfrak{n}] \backslash \bar{I}} \eta_i \ge t\right\} \le \frac{\left(1 + e\mathfrak{p}|I|\right)^{|[\mathfrak{n}] \backslash \bar{I}|}}{\exp(\lambda t)}, \qquad t > 0.$$

In particular, taking $t := \frac{\delta_0}{32} np|I|$ and $\lambda := \log \frac{1}{\mathfrak{p}|I|}$, we get

$$\begin{aligned} \mathbb{P}\left\{\sum_{i \in [\mathfrak{n}] \backslash \bar{I}} \eta_i \ge \frac{\delta_0}{32} np|I|\right\} &\le \exp\left(e\mathfrak{p}\mathfrak{n}|I| - \lambda \frac{\delta_0}{32} np|I|\right) \le \exp\left(-\lambda \frac{\delta_0}{64} np|I|\right) \\ &\le \exp\left(-\log\left(\frac{1}{2c_{3.7}}\right) \cdot \frac{\delta_0}{64} np|I|\right) \le n^{-2|I|}, \end{aligned} \tag{A.10}$$

where the second and the third inequalities follow from recalling that $p|I| \le c_{3.7}$ for some sufficiently small constant $c_{3.7}$, depending only on δ_0, and the last inequality follows from our assumption that $np \ge \log n/2$ and shrinking $c_{3.7}$ even further, if necessary.

To complete the proof of the fact that (5) holds with high probability, we show that

$$\mathbb{P}\left(\sum_{j \in I} \left|\text{supp}(\text{col}_j(\text{fold}(A_n))) \cap \bar{I}\right| \ge \frac{\delta_0}{32} np|I|\right) \le 2n^{-2|I|}. \tag{A.11}$$

Now the proof finishes from (A.10)-(A.11) by first taking an union over $I \in \binom{[n]}{k}$ followed by a union over $k = 2, 3, \ldots, c_{3.7} p^{-1}$. We omit the details.

Turning to prove (A.11) we we denote $\hat{I}(I) := \hat{I} := \cup_{i \in \bar{I}} \{i, \mathfrak{n} + i\}$. As the entries of A_n are $\{0,1\}$-valued, we see that

$$\text{supp}(\text{col}_j(\text{fold}(A_n))) \cap \bar{I} \subset \text{supp}(\text{col}_j(A_n)) \cap \hat{I}.$$

Moreover, $I \subset \hat{I}$. Therefore, it is enough to show that

$$\mathbb{P}\left(\sum_{j \in \hat{I}} \left|\text{supp}(\text{col}_j(A_n)) \cap \hat{I}\right| \ge \frac{\delta_0}{32} np|I|\right) \le 2n^{-2|I|}. \tag{A.12}$$

Since A_n satisfies Assumption 3.1 we have that the upper triangular part of the sub-matrix of A_n induced by the rows and columns indexed by $\hat{I}$ consists of independent $\{0,1\}$-valued random variables stochastically dominated by i.i.d. $\text{Ber}(p)$ variables. So does the lower triangular part of that sub-matrix.

For ease of writing let us write

$$\mathscr{X}_L := \sum_{i \ge j \in \hat{I}} a_{i,j} \quad \text{and} \quad \mathscr{X}_U := \sum_{i \le j \in \hat{I}} a_{i,j}$$

and note $\mathscr{X}_U$ and $\mathscr{X}_L$ has the same law. Thus to establish (A.12) it suffices to show that

$$\mathbb{P}(\mathscr{X}_U \ge \frac{\delta_0}{64} np|I|) \le n^{-2|I|}. \tag{A.13}$$

The above is obtained by using the Laplace transform method as above. Indeed, we note that

$$\mathbb{P}(\mathscr{X}_U = \ell) \le \binom{|\hat{I}|^2}{\ell} p^\ell, \quad \ell \in \mathbb{N} \cup \{0\}$$

and therefore

$$\mathbb{E}\left[\exp\left(\lambda X_U\right)\right] \le \exp\left(e^\lambda p |\hat{I}|^2\right) \le \exp(4|I|),$$

where $\lambda = \log \frac{1}{p|I|}$ and we have used the fact that $|\hat{I}| \le 2|I|$. Hence, upon using Markov's inequality and proceeding similarly as in (A.10) we deduce (A.13). It completes the proof of (A.12).

Now it remains to prove that property (6) holds with high probability. Recalling the definition of the folded matrix again we note that $|\operatorname{supp}(\operatorname{col}_j(\operatorname{fold}(A_n)))| \le |\operatorname{supp}(\operatorname{col}_j(A_n))|$. To show that the cardinality of the support of $\operatorname{col}_j(\operatorname{fold}(A_n))$ is not too small compared to its unfolded version we observe that if $k \in \operatorname{supp}(\operatorname{col}_j(A_n))$ but $k \notin \operatorname{supp}(\operatorname{col}_j(\operatorname{fold}(A_n)))$ then we must have that $a_{k,j} = a_{k,\mathfrak{n}+j} = 1$. Using estimates on the binomial probability and Chernoff bound we show that number of such k is small.

To carry out the above heuristic, we fix $j \in [n]$ and since the entries of A_n are $\{0,1\}$ valued we note that

$$|\operatorname{supp}(\operatorname{col}_j(\operatorname{fold}(A_n)))| = \sum_{i \in [\mathfrak{n}]} \left[a_{i,j} \cdot (1 - a_{i+\mathfrak{n},j}) + a_{i+\mathfrak{n},j} \cdot (1 - a_{i,j})\right]$$

Further observe that

$$|\operatorname{supp}(\operatorname{col}_j(A_n)) \cap [\mathfrak{n}]| = \sum_{i=1}^{\mathfrak{n}} a_{i,j} = \sum_{i=1}^{\mathfrak{n}} a_{i,j} \cdot a_{i+\mathfrak{n},j} + \sum_{i=1}^{\mathfrak{n}} a_{i,j} \cdot (1 - a_{i+\mathfrak{n},j})$$

and

$$|\operatorname{supp}(\operatorname{col}_j(A_n)) \cap ([2\mathfrak{n}] \backslash [\mathfrak{n}])| = \sum_{i=\mathfrak{n}+1}^{2\mathfrak{n}} a_{i,j} = \sum_{i=1}^{\mathfrak{n}} a_{i,j} \cdot a_{i+\mathfrak{n},j} + \sum_{i=1}^{\mathfrak{n}} a_{i+\mathfrak{n},j} \cdot (1 - a_{i,j}).$$

Therefore,

$$||\operatorname{supp}(\operatorname{col}_j(A_n))| - |\operatorname{supp}(\operatorname{col}_j(\operatorname{fold}(A_n)))|| \le 2 \sum_{i=1}^{\mathfrak{n}} a_{i,j} \cdot a_{i+\mathfrak{n},j} + 1.$$

Denoting

$$\Delta_j := \sum_{i=1}^{\mathfrak{n}} a_{i,j} \cdot a_{i+\mathfrak{n},j},$$

we see that Δ_j is stochastically dominated by $\operatorname{Bin}(\mathfrak{n}, p^2)$. To finish the proof we need to find bounds on Δ_j.

First let us consider the case $p \le n^{-5/12}$. For any $k_0 \in \mathbb{N}$, sufficiently large, we see that

$$\mathbb{P}(\Delta_j \ge k_0) \le \binom{\mathfrak{n}}{k_0} p^{2k_0} \le (np^2)^{k_0} \le n^{-k_0/6} \le n^{-2}. \tag{A.14}$$

For $n^{-5/12} \le p \le c$, for some small $c > 0$ depending on δ_0, we use Chernoff bound to deduce that

$$\mathbb{P}\left(\Delta_j \ge \frac{\delta_0}{16} np\right) \le \mathbb{P}\left(\Delta_j \ge 2p^{-1/2} \cdot \mathfrak{n}p^2\right) \le \exp\left(-\frac{1}{3}p^{-1/2} \cdot \mathfrak{n}p^2\right) \le \exp\left(-\frac{1}{9}n^{3/8}\right). \tag{A.15}$$

Combining (A.14)-(A.15) and taking an union over $j \in [n]$ we show that property (6) holds with high probability. This completes the proof of the lemma. □

Appendix B. Proof of invertibility over sparse vectors with a large spread component

In this section we prove Proposition 3.18. As already mentioned in Section 3.3 the proof is similar to that of Proposition 3.15. There are two key differences. Since our goal is to find a uniform bound on $\|A_n x\|_2$ for x's with a large spread component, unlike in the proof of Proposition 3.18, we use Lemma 3.19 to estimate the small ball probability. Moreover, as noted earlier, Assumption 3.1 allows some dependencies among its entries. Therefore, to tensorize the small ball probability we need to extract a sub-matrix of A_n with jointly independent entries such that the coordinates of x corresponding to the columns of this chosen sub-matrix form a vector with a large spread component and a sufficiently large norm. Below we make this idea precise.

Proof of Proposition 3.18. First, let us show that (3.40) implies (3.42). To this end, we begin by noting that if $c_{3.18} < \frac{1}{2}$ then for any $x \in \text{Dom}(c_0^* n, c_{3.18} K^{-1})$ we have that $\|x_{[M_0+1:c_0^* n]}\|_2 \ge \|x_{[c_0^* n+1:n]}\|_2$ (see also (3.33)). Hence, for $x \notin V_{M_0}$ we obtain that $\|x_{[M_0+1:c_0^* n]}\|_2 \ge \rho/\sqrt{2}$. Therefore, (3.40) implies that

$$\mathbb{P}\left(\{\exists x \in V_{c_0^*, c_{3.18}} \backslash V_{M_0} : \|(A_n - p\boldsymbol{J}_n)x - y\|_2 \le 2\widetilde{c}_{3.18}\rho\sqrt{np}\} \cap \Omega_K^0\right) \le \exp(-2\bar{c}_{3.18} n),$$

where we recall the definition of Ω_K^0 from (3.28). Hence, proceeding as in the steps leading to (3.30) we deduce (3.42) upon assuming (3.40).

So, to complete the proof of the proposition it remains to establish (3.40). To prove it, we fix $x \notin V_{M_0}$. Then

$$\|x_{[M_0+1:n]}\|_2 \ge \rho \quad \text{and} \quad \frac{\|x_{[M_0+1:n]}\|_\infty}{\|x_{[M_0+1:n]}\|_2} \le \frac{K}{c_{3.15}} \cdot \sqrt{\frac{\log n}{n\sqrt{\log\log n}}}.$$

Fixing $\bar{c}_0 \in (c_0^*, 1)$, as $M_0 \le \frac{1-\bar{c}_0}{2} n$ for all large n, recalling the fact that the non-zero entries of $x_{[m_1:m_2]}$, for $m_1 < m_2$, are the coordinates of x that take places from m_1 to m_2 in the non-increasing arrangement according their absolute values, we note that

$$\|x_{[M_0+1:n]}\|_2^2 = \|x_{[M_0+1:(1-\bar{c}_0)n]}\|_2^2 + \|x_{[(1-\bar{c}_0)n+1:n]}\|_2^2 \le \frac{1+\bar{c}_0}{1-\bar{c}_0} \cdot \|x_{[M_0+1:(1-\bar{c}_0)n]}\|_2^2. \tag{B.1}$$

Therefore

$$\|x_{[M_0+1:(1-\bar{c}_0)n]}\|_2 \ge \rho \cdot \sqrt{\frac{1-\bar{c}_0}{1+\bar{c}_0}} \quad \text{and} \quad \frac{\|x_{[M_0+1:(1-\bar{c}_0)n]}\|_\infty}{\|x_{[M_0+1:(1-\bar{c}_0)n]}\|_2} \le \frac{K}{c_{3.15}} \cdot \sqrt{\frac{1+\bar{c}_0}{1-\bar{c}_0}} \cdot \sqrt{\frac{\log n}{n\sqrt{\log\log n}}}.$$

Note that this shows $x_{[M_0+1:(1-\bar{c}_0)n]}$ has a large spread part and a large norm. Denoting $\mathcal{I} := \mathcal{I}(x) := \text{supp}(x_{[M_0+1:(1-\bar{c}_0)n]})$ we note that Assumption 3.1 implies that the entries $\{a_{i,j}\}_{j\in\mathcal{I}, i\notin\mathcal{I}}$ are i.i.d. Ber(p). So, now we can carry out the scheme that was outlined above by using the joint independence of $\{a_{i,j}\}_{j\in\mathcal{I}, i\notin\mathcal{I}}$.

Indeed, using Lemma 3.19 we find that for any $i \notin \mathcal{I}$, $y \in \mathbb{R}^n$, and $\varepsilon_0 > 0$ we have

$$\begin{aligned} \text{(B.2)} \quad & \mathbb{P}\left(|((A_n - p\boldsymbol{J}_n)x)_i - y_i| \le p^{1/2}(1-p)^{1/2} \|x_{[M_0+1:(1-\bar{c}_0)n]}\|_2 \cdot \bar{c}_0^{-1/2}\varepsilon_0 \right) \\ & \le \mathcal{L}\left(\langle \boldsymbol{a}_i, x\rangle, p^{1/2}(1-p)^{1/2} \|x_{[M_0+1:(1-\bar{c}_0)n]}\|_2 \cdot \bar{c}_0^{-1/2}\varepsilon_0 \right) \\ & \le C_{3.19} \left(\frac{\varepsilon_0}{\sqrt{\bar{c}_0}} + \frac{2K}{c_{3.15}} \cdot \sqrt{\frac{1+\bar{c}_0}{1-\bar{c}_0}} \cdot \sqrt{\frac{\log n}{np\sqrt{\log\log n}}} \right) \le 2C_{3.19} \frac{\varepsilon_0}{\sqrt{\bar{c}_0}}, \end{aligned}$$

for all sufficiently large n (depending only on ε_0), where $\boldsymbol{a}_i$ is the i-th row of A_n and we have used the fact that $np \ge c_1 \log n$ for some $c_1 > 0$. We will choose ε_0 as a small constant during the course of the proof.

Since the entries $\{a_{i,j}\}_{j\in\mathcal{I}, i\notin\mathcal{I}}$ are i.i.d. Ber(p), we apply a standard tensorization argument, for example [34, Lemma 5.4], to deduce from (B.2) that for any $x \notin V_{M_0}$

$$\begin{aligned} \text{(B.3)} \quad & \mathbb{P}\left(\|(A_n - p\boldsymbol{J}_n)x - y\|_2 \le \sqrt{np(1-p)} \|x_{[M_0+1:(1-\bar{c}_0)n]}\|_2 \varepsilon_0 \right) \\ & \le \mathbb{P}\left(\sum_{i\notin\mathcal{I}} |((A_n - p\boldsymbol{J}_n)x)_i - y_i|^2 \le p(1-p)\|x_{[M_0+1:(1-\bar{c}_0)n]}\|_2^2 \varepsilon_0^2 \bar{c}_0^{-1} \cdot |\mathcal{I}^c| \right) \le (C_0 \cdot \varepsilon_0)^{\bar{c}_0 n}, \end{aligned}$$

for some constant C_0, depending only on $\bar{c}_0$, where the last two steps follow from the fact that $|\mathcal{I}^c| \ge \bar{c}_0 n$ and upon choosing ε_0 such that $C_0 \cdot \varepsilon_0 \le \frac{1}{2}$.

To complete the proof we use an ε-net similar to the proof of Proposition 3.15. First, setting

$$\varepsilon = \frac{\rho}{2}\tau = \frac{\varepsilon_0 \rho}{448K} \cdot \sqrt{\frac{1-\bar{c}_0}{1+\bar{c}_0}}, \tag{B.4}$$

and using Fact 3.17 we obtain a net $\widetilde{\mathcal{M}}$ in $V_{c_0^*, c_{3.18}} \backslash V_{M_0}$ with

$$|\widetilde{\mathcal{M}}| \le \bar{C}^n \binom{n}{M_0} \binom{n}{c_0^* n} \left(\frac{1}{\varepsilon_0}\right)^{c_0^* n+1} \left(\frac{1}{\rho}\right)^{M_0+1} \le \bar{C}^n \left(\frac{en}{M_0}\right)^{M_0} \left(\frac{e}{c_0^*}\right)^{c_0^* n} \left(\frac{1}{\varepsilon_0}\right)^{c_0^* n+1} \left(\frac{1}{\rho}\right)^{M_0+1},$$

for some $\bar{C}$, depending only on $\bar{c}_0$ and c_0^*. Recalling that $M_0 = \frac{n\sqrt{\log\log n}}{\log n}$ and the definition of ρ we observe that

$$\left(\frac{n}{M_0 \rho^2}\right)^{M_0} = \exp(o(n)),$$

for $p \in (0,1)$ satisfying $np \ge c_1 \log n$. Therefore, we further have that

$$|\widetilde{\mathcal{M}}| \le C_\star^n \cdot \left(\frac{1}{\varepsilon_0}\right)^{c_0^* n+1}, \tag{B.5}$$

for some other constant $C_\star$, depending only on c_0^* and $\bar{c}_0$. Next proceeding as in the steps leading to (3.36) we obtain that for any $x \in V_{c_0^*, c_{3.18}} \backslash V_{M_0}$ there exists $\bar{x} \in \widetilde{\mathcal{M}}$ such that for any $y \in \mathbb{R}^n$

$$\|(A_n - p\boldsymbol{J}_n)\bar{x} - y\|_2 \le \|(A_n - p\boldsymbol{J}_n)x - y\|_2 + 4K\sqrt{np} \cdot \varepsilon + 2K\sqrt{np} \cdot \tau \cdot \|v_{\bar{x}}\|_2 + 12c_{3.18}\sqrt{np} \cdot \|v_{\bar{x}}\|_2 .$$

Since $\|\bar{x}_{[M_0+1:c_0^* n]}\|_2 = \|v_{\bar{x}}\|_2 \ge \rho/\sqrt{2}$, using (B.4) and setting

$$c_{3.18} \le \frac{\varepsilon_0}{56} \cdot \sqrt{\frac{1-\bar{c}_0}{1+\bar{c}_0}}, \tag{B.6}$$

we deduce from above that any $x \in V_{c_0^*, c_{3.18}} \backslash V_{M_0}$ there exists $\bar{x} \in \widetilde{\mathcal{M}}$ such that for any $y \in \mathbb{R}^n$

$$\|(A_n - p\boldsymbol{J}_n)\bar{x} - y\|_2 \le \|(A_n - p\boldsymbol{J}_n)x - y\|_2 + \frac{\varepsilon_0}{7} \cdot \sqrt{\frac{1-\bar{c}_0}{1+\bar{c}_0}} \|\bar{x}_{[M_0+1:c_0^* n]}\|_2 \cdot \sqrt{np}.$$

Furthermore, by our construction of the net $\widetilde{\mathcal{M}}$,

$$\|x_{[M_0+1:c_0^* n]}\|_2 \le \|\bar{x}_{[M_0+1:c_0^* n]}\|_2 + \varepsilon \le \left(1 + \frac{\varepsilon_0}{224K}\right) \cdot \|\bar{x}_{[M_0+1:c_0^* n]}\|_2.$$

Therefore, upon assuming $p \le \frac{1}{4}$ and recalling (B.1), this further yields that

$$\begin{aligned}
\text{(B.7)} \quad & \mathbb{P}\left(\exists x \in V_{c_0^*, c_{3.18}} \backslash V_{M_0} : \|(A_n - p\boldsymbol{J}_n)x - y\|_2 \le \frac{\varepsilon_0}{4} \|x_{[M_0+1:c_0^* n]}\|_2 \cdot \sqrt{\frac{1-\bar{c}_0}{1+\bar{c}_0}} \cdot \sqrt{np}\right) \\
& \le \mathbb{P}\left(\exists \bar{x} \in \widetilde{\mathcal{M}} : \|(A_n - p\boldsymbol{J}_n)\bar{x} - y\|_2 \le \sqrt{np(1-p)} \|\bar{x}_{[M_0+1:(1-\bar{c}_0)n]}\|_2 \varepsilon_0\right) \\
& \le |\widetilde{\mathcal{M}}| \cdot (C_0 \cdot \varepsilon_0)^{\bar{c}_0 n} \le C_0^{\bar{c}_0 n} C_\star^n \varepsilon_0^{-1} \varepsilon_0^{(\bar{c}_0 - c_0^*)n} \le \varepsilon_0^{\frac{\bar{c}_0 - c_0^*}{2} n},
\end{aligned}$$

where the second last step follows from (B.5) and the last step follows upon using the fact that $\bar{c}_0 > c_0^*$ and choosing ε_0 sufficiently small. This yields (3.40) and hence the proof of the proposition is complete. □

References

[1] Z. D. Bai, and J. W. Silverstein. *Spectral analysis of large dimensional random matrices*. 2nd ed. Springer Series in Statistics. Dordrecht: Springer, 2010.

[2] A. S. Banderia and R. van Handel. Shrap nonasymptotic bounds on the norm of random matrices with independent entries. *The Annals of Probability*, 44(4), 2479–2506, 2016.

[3] A. Basak, N. Cook, and O. Zeitouni. Circular law for the sum of random permutations. *Elec. J. Probab.*, paper no. 33, 51 pp, 2018.

[4] A. Basak and A. Dembo. Limiting spectral distribution of sum of unitary and orthogonal matrices. *Elec. Comm. Probab.*, article 69, 2013

[5] A. Basak and M. Rudelson. Invertibility of sparse non-Hermitian matrices. *Advances in Mathematics*, 310, 426–483, 2017.

[6] A. Basak and M. Rudelson. The circular law for sparse non-Hermitian matrices. *Ann. Probab.* to appear. Available at https://arxiv.org/abs/1707.03675.

[7] C. Bordenave, P. Caputo, and D. Chafaï. Circular Law Theorem for Random Markov Matrices. *Prob. Th. Rel. Field.*, 152 (3-4), 751-779, 2012.

[8] C. Bordenave and D. Chafaï. Around the circular law. *Probability Surveys*, 9(0), 1-89, 2012.

[9] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities: A nonasymptotic theory of independence.* Oxford university press, 2013.

[10] J. Bourgain, V. Vu, and P. M. Wood. On the singularity probability of discrete random matrices. *J. Funct. Anal.*, 258(2), 559–603, 2010.

[11] N. Cook. On the singularity of adjacency matrices for random regular digraphs. *Probab. Theo.Rel. Flds.*, 167(1–2), 143–200, 2017.

[12] N. Cook. The circular law for random regular digraphs. ArXiv preprint arXiv:1703.05839v2, 2017.

[13] K. P. Costello. Bilinear and quadratic variants on the Littlewood-Offord problem. *Israel Journal of Mathematics*, 194(1), 359–394, 2013.

[14] K. P. Costello and V. H. Vu. The rank of random graphs. *Rand. Struct. & Algo.*, 33(3), 269–285, 2008.

[15] K. P. Costello and V. Vu. On the rank of random sparse matrices. *Combin., Prob. & Comput.*, 19(3), 321–342, 2010.

[16] A. Edelman. Eigenvalues and condition numbers of random matrices. *SIAM J. Matrix Anal. Appl.*, 9, 543–560, 1988.

[17] A. Frieze. Random structures and algorithms. In *Proceedings of the International Congress of Mathematicians – Seoul 2014*. Vol. 1, pages 311–340. Kyung Moon Sa, Seoul, 2014.

[18] J. Huang. Invertibility of adjacency matrices for random d-regular directed graphs. *ArXiv preprint* arXiv:1806.01382v2, 2018.

[19] J. Huang. Invertibility of adjacency matrices for random d-regular graphs. *ArXiv preprint* arXiv:1807.06465v1, 2018.

[20] F. Götze and A. Tikhomirov. The circular law for random matrices. *Ann. Probab.*, 38 (4), 1444-1491, 2010.

[21] J. Kahn, J. Komlós, and E. Szemerédi. On the probability that a random ± 1 matrix is singular. *J. Amer. Math. Soc.*, 8(1), 223–240, 1995.

[22] J. Komlós. On the determinant of $(0,1)$ matrices. *Studia. Sci. Math. Hungar.* 2, 7–22, 1967.

[23] J. Komlós. On the determinant of random matrices. *Studia. Sci. Math. Hungar.* 3, 387–399, 1968.

[24] J. Komlós. Circulated manuscript. Edited version available online at http://sites.math.rutgers.edu/~komlos/01short.pdf, 1977.

[25] B. Landon, P. Sosoe, and H.-T. Yau. Fixed energy universality of Dyson Brownian motion. *ArXiv preprint* arXiv:1609.09011v2, 2016.

[26] R. Latała. Some estimates of norms of random matrices. *Proc. Amer. Math. Soc.*, 133(5), 1273–1282, 2004.

[27] A. E. Litvak, A. Lytova, K. Tikhomirov, N. Tomczak-Jaegermann, and P. Youssef. Adjacency matrices of random digraphs: singularity and anti-concentration. *J. Math. Anal. Appl.*, 445(2):1447–1491, 2017.

[28] A. E. Litvak, A. Lytova, K. Tikhomirov, N. Tomczak-Jaegermann, and P. Youssef. The smallest singular value of a shifted d-regular random square matrix. *Probab. Thero. Rel. Flds.*, to appear. Preprint available at https://arxiv.org/abs/1707.02635.

[29] A. E. Litvak, A. Lytova, K. Tikhomirov, N. Tomczak-Jaegermann, and P. Youssef. The rank of random regular digraphs of constant degree. *ArXiv preprint* arXiv:1801.05577, 2018.

[30] A. E. Litvak, A. Lytova, K. Tikhomirov, N. Tomczak-Jaegermann, and P. Youssef. Circular law for sparse random regular digraphs. *ArXiv preprint* arXiv:1801.05576v2, 2018.

[31] A. E. Litvak, A. Pajor, M. Rudelson, and N. Tomczak-Jaegermann. Smallest singular value of random matrices and geometry of random polytopes. *Adv. Math.* 195, 491–523, 2005.

[32] A. Mészáros. The distribution of sandpile groups of random regular graphs. *ArXiv preprint* arXiv:1806.03736v1, 2018.

[33] H. H. Nguyen and W. M. Wood. Cokernels of adjacency matrices of random r-regular graphs. *ArXiv preprint* arXiv:1806.10068v2, 2018.

[34] M. Rudelson. Invertibility of random matrices: Norm of the inverse. *Ann. of Math.*, 168, 575-600, 2008.

[35] M. Rudelson and K. Tikhomirov. The sparse circular law under minimal assumptions. *ArXiv preprint* arXiv:1807.08085v1, 2018.

[36] M. Rudelson and R. Vershynin. The Littlewood-Offord Problem and invertibility of random matrices. *Adv. Math.* 218 (2), 600–633, 2008.

[37] M. Rudelson and R. Vershynin. Smallest singular value of a random rectangular matrix. *Comm. Pur. App. Math.*, 62, 1707-1739, 2009.

[38] M. Rudelson and R. Vershynin. Invertibility of random matrices: unitary and orthogonal perturbations. *J. Amer. Math. Soc.*, 27(2), 293–338, 2014.

[39] S. Smale. On the efficiency of algorithms of analysis. *Bull. Amer. Math. Soc. (N.S.)*, 13, 87–121, 1985.

[40] T. Tao. Least singular value, circular law, and Lindeberg exchange. *Preprint*, available at http://helper.ipam.ucla.edu/publications/qlatut/qlatut_15156.pdf, 2017.

[41] T. Tao and V. Vu. On random ± 1 matrices: singularity and determinant. *Rand. Struct. & Algo.*, 28, 1–23, 2006.

[42] T. Tao and V. Vu. On the singularity probability of random Bernoulli matrices. *J. Amer. Math. Soc.*, 20, 603–628, 2007.

[43] T. Tao and V Vu. Random matrices: the circular law. *Comm. Contem. Math.*, 10(2), 261–307, 2008.

[44] T. Tao and V. Vu. Random matrices: universality of the ESDs and the circular law. *Ann. Probab.*, 38, 2023-2065, 2010 (with an appendix by M. Krishnapur).

[45] R. Vershynin. Invertibility of symmetric random matrices. *Random Structures and Algorithms*, 44(2), 135–182, 2014.

[46] J. von Neumann. *Collected works. Vol. V: Design of computers, theory of automata and numerical analysis.* General editor: A. H. Taub. A Pergamon Press Book The Macmillan Co., New York 1963.

[47] J. von Neumann and H. H. Goldstine. Numerical inverting of matrices of high order. *Bull. Amer. Math. Soc.*, 53, 1021–1099, 1947.

[48] V. Vu. Random discrete matrices. In *Horizons of combinatorics*, Vol. 17 of *Bolyai Soc. Math. Stud.*, 257–280. Springer, Berlin, 2008.

[49] V. Vu. Combinatorial problems in random matrix theory. In *Proceedings of the International Congress of Mathematicians – Seoul 2014*. Vol. 4, 489–508, Kyung Moon Sa, Seoul, 2014.
[50] F. Wei. Investigate invertibility of sparse symmetric matrices. *ArXiv preprint* arXiv:1712.04341v2, 2017.
[51] P. M. Wood. Universality and the circular law for sparse random matrices. *Ann. Appl. Probab.*, 22(3), 1266–1300, 2012.

*International Center for Theoretical Sciences
Tata Institute of Fundamental Research
Bangalore 560089, India
and
Department of Mathematics, Weizmann Institute of Science
POB 26, Rehovot 76100, Israel

† Department of Mathematics, University of Michigan
East Hall, 530 Church Street
Ann Arbor, Michigan 48109, USA