arXiv: 1610.01765

The spectral gap of dense random regular graphs

Konstantin Tikhomirov

Department of Mathematics,
Princeton University
e-mail: kt12@math.princeton.edu

and

Pierre Youssef

Laboratoire de Probabilités et de Modèles aléatoires,
Université Paris Diderot
e-mail: youssef@math.univ-paris-diderot.fr

Abstract: For any $\alpha \in (0,1)$ and any $n^\alpha \leq d \leq n/2$, we show that $\lambda(\mathbf{G}) \leq C_\alpha\sqrt{d}$ with probability at least $1-\frac{1}{n}$, where $\mathbf{G}$ is the uniform random undirected d-regular graph on n vertices, $\lambda(\mathbf{G})$ denotes its second largest eigenvalue (in absolute value) and C_α is a constant depending only on α. Combined with earlier results in this direction covering the case of sparse random graphs, this completely settles the problem of estimating the magnitude of $\lambda(\mathbf{G})$, up to a multiplicative constant, for all values of n and d, confirming a conjecture of Vu. The result is obtained as a consequence of an estimate for the second largest singular value of adjacency matrices of random *directed* graphs with predefined degree sequences. As the main technical tool, we prove a concentration inequality for arbitrary linear forms on the space of matrices, where the probability measure is induced by the adjacency matrix of a random directed graph with prescribed degree sequences. The proof is a non-trivial application of the Freedman inequality for martingales, combined with self-bounding and tensorization arguments. Our method bears considerable differences compared to the approach used by Broder, Frieze, Suen and Upfal (1999) who established the upper bound for $\lambda(\mathbf{G})$ for $d = o(\sqrt{n})$, and to the argument of Cook, Goldstein and Johnson (2015) who derived a concentration inequality for linear forms and estimated $\lambda(\mathbf{G})$ in the range $d = O(n^{2/3})$ using size-biased couplings.

MSC 2010 subject classifications: Primary 05C80; secondary 60B20.
Keywords and phrases: random regular graphs, uniform model, spectral gap, random matrices.

1. Introduction

Let n be a natural number and let $d \leq n$. An *undirected d-regular* graph G with the vertex set $\{1, 2, \dots, n\}$ is a graph in which every vertex has exactly d neighbors. Spectral properties of random undirected d-regular graphs have attracted considerable attention of researchers. Regarding the empirical spectral distribution, we refer, among others, to a classical result of McKay [21], as well as more recent papers [14, 27, 4]. A new line of research deals with invertibility of adjacency matrices [11, 19]. The seminal works of Alon and Milman [2] and Alon [1] established a connection between the magnitude of the second largest eigenvalue of a regular graph with its expansion properties. The conjecture of Alon [1] on the limit of the spectral gap when the degree is fixed and the number of vertices tends to infinity, was resolved by Friedman [16] (see [9, 18, 17] for earlier results). Friedman proved, in particular, that $\lambda(\mathbf{G}) = 2\sqrt{d-1} + o(1)$ with probability tending to one with $n \to \infty$, where $\lambda(\mathbf{G})$ is the second largest (in absolute value) eigenvalue of the undirected d-regular random graph $\mathbf{G}$ on n vertices, uniformly distributed on the set of all simple d-regular graphs (see [7] for an alternative proof of Friedman's theorem; see also [22] for a different approach producing a weaker bound). A natural extension of Alon's question to the setting when d grows with n to infinity, was considered in [8, 13, 12]. Namely, in [8] the authors showed that for $d = o(\sqrt{n})$, one has $\lambda(\mathbf{G}) \le C\sqrt{d}$ with probability tending to one with n for some universal constant $C > 0$. This result was extended to the range $d = O(n^{2/3})$ in [12]. In [13], the bound $\lambda(\mathbf{G}) \le C\sqrt{d}$ w.h.p. was obtained for $\mathbf{G}$ distributed according to *the permutation model*, which we do not consider here.

In [28], Vu conjectured that $\lambda(\mathbf{G}) = (2 + o(1))\sqrt{d - d^2/n}$ w.h.p. in the uniform model, when $d \le n/2$ and d tends to infinity with n (see also [29, Conjectures 7.3, 7.4]). The "isomorphic" version of this question was one of the motivations for our work. Apart from the previously mentioned connection with structural properties of random graphs, this line of research seems quite important in another aspect as well. Random d-regular graphs supply a natural model of randomness for square matrices, in which the matrix cannot be partitioned into independent disjoint blocks (say, rows or columns) but the correlation between *very small* disjoint blocks is weak. Techniques developed to deal with the adjacency matrices of these graphs may prove useful in other problems within the random matrix theory. In this respect, our intention was to develop, or rely on, arguments which are flexible and admit various generalizations.

Given an $n \times n$ symmetric matrix A, we let $\lambda_1(A) \ge \lambda_2(A) \ge \ldots \ge \lambda_n(A)$ be its eigenvalues arranged in non-increasing order (counting multiplicities). For an undirected graph G on n vertices, we define $\lambda_1(G), \ldots, \lambda_n(G)$ as the eigenvalues of its adjacency matrix.

Theorem A. *For every $\alpha \in (0,1)$ and $m \in \mathbb{N}$ there are $L = L(\alpha, m) > 0$ and $n_0 = n_0(\alpha, m)$ with the following property: Let $n \ge n_0$, $n^\alpha \le d \le n/2$, and let $\mathbf{G}$ be a random graph uniformly distributed on the set $\mathcal{G}_n(d)$ of simple undirected d-regular graphs on n vertices. Then*

$$\max\big(|\lambda_2(\mathbf{G})|, |\lambda_n(\mathbf{G})|\big) \le L\sqrt{d}$$

with probability at least $1 - n^{-m}$.

Note that, combined with [8, 12], our theorem gives $\max\big(|\lambda_2(\mathbf{G})|, |\lambda_n(\mathbf{G})|\big) = O(\sqrt{d})$ w.h.p. for all $d \le n/2$. Denote by $\mathbf{M}$ the adjacency matrix of $\mathbf{G}$. It is easy to see that (deterministically) d is the largest eigenvalue of $\mathbf{M}$ with $\mathbf{1}$ (vector of ones) as the corresponding eigenvector. Hence, from the Courant–Fischer formula, we obtain $\lambda(\mathbf{G}) = \|\mathbf{M} - \frac{d}{n}\mathbf{1}\,\mathbf{1}^t\| = \|\mathbf{M} - \mathbb{E}\mathbf{M}\|$. Theorem A thus implies that the spectral measure of $\frac{1}{\sqrt{d}}(\mathbf{M} - \mathbb{E}\mathbf{M})$ is supported on an interval of constant length with probability going to one with $n \to \infty$. We refer to [4, 3] (and references therein) for recent advances concerning the limiting behavior of the spectral measure for random d-regular graphs in the uniform model. The proof of Theorem A is obtained by a rather general yet simple procedure which reduces the question to the non-symmetric (i.e. directed) setting, which we are about to consider.

A *directed d-regular* graph G on n vertices is a directed (labeled) graph in which every vertex has d in-neighbors and d out-neighbors. We allow directed graphs to have loops, but do not allow multiple edges (edges connecting the same pair of vertices in opposite directions are distinct). The corresponding set of graphs will be denoted by $\mathcal{D}_n(d)$. Note that the set of adjacency matrices for graphs in $\mathcal{D}_n(d)$ is the set of all 0-1-matrices with the sum of elements in each row and column equal to d. Note also that there is a natural bijection from $\mathcal{D}_n(d)$ onto the set of *bipartite* d-regular simple undirected graphs on $2n$ vertices. Given an $n \times n$ matrix A, we let $s_1(A) \geq s_2(A) \geq \ldots \geq s_n(A)$ be its singular values arranged in non-increasing order (counting multiplicities). For a directed graph G on n vertices, we define $s_1(G), \ldots, s_n(G)$ as the singular values of its adjacency matrix.

Theorem B. *For every $\alpha \in (0,1)$ and $m \in \mathbb{N}$ there are $L = L(\alpha, m) > 0$ and $n_0 = n_0(\alpha, m) \in \mathbb{N}$ with the following property: Let $n \geq n_0$, and let $n^\alpha \leq d \leq n/2$. Further, let $\mathbf{G}$ be a random directed d-regular graph uniformly distributed on $\mathcal{D}_n(d)$. Then*

$$s_2(\mathbf{G}) \leq L\sqrt{d}$$

with probability at least $1 - n^{-m}$. Consequently, if $\widetilde{\mathbf{G}}$ is a random undirected graph uniformly distributed on the set of all bipartite simple d-regular graphs on $2n$ vertices then

$$\lambda_2(\widetilde{\mathbf{G}}) \leq L\sqrt{d}$$

with probability at least $1 - n^{-m}$.

Theorem B above is stated for reader's convenience. In fact, we prove a more general statement which deals with random graphs with *predefined degree sequences*. With every directed graph G on $\{1, 2, \ldots, n\}$, we associate two degree sequences: *the in-degree sequence* $\mathbf{d}^{in}(G) = (\mathbf{d}^{in}_1, \mathbf{d}^{in}_2, \ldots, \mathbf{d}^{in}_n)$, with $\mathbf{d}^{in}_i$ equal to the number of in-neighbors of vertex i, and *the out-degree sequence* $\mathbf{d}^{out}(G) = (\mathbf{d}^{out}_1, \mathbf{d}^{out}_2, \ldots, \mathbf{d}^{out}_n)$, where $\mathbf{d}^{out}_i$ is the number of out-neighbors of i ($i \leq n$). Conversely, given two integer vectors $\mathbf{d}^{in}, \mathbf{d}^{out} \in \mathbb{R}^n$, we will denote by $\mathcal{D}_n(\mathbf{d}^{in}, \mathbf{d}^{out})$ the set of all directed graphs on n vertices with the in- and out-degree sequence $\mathbf{d}^{in}$ and $\mathbf{d}^{out}$, respectively. Again, we allow the graphs to have loops but do not allow multiple edges.

Let us introduce the following two Orlicz norms in $\mathbb{R}^n$:

$$\|x\|_{\psi,n} := \inf\Big\{\lambda > 0 : \frac{1}{en}\sum_{i=1}^n e^{|x_i|/\lambda} \le 1\Big\}, \qquad x = (x_1,\dots,x_n) \in \mathbb{R}^n; \tag{1}$$

$$\|x\|_{\log,n} := \inf\Big\{\lambda > 0 : \frac{1}{n}\sum_{i=1}^n \frac{|x_i|}{\lambda}\ln_+\Big(\frac{|x_i|}{\lambda}\Big) \le 1\Big\}, \qquad x = (x_1,\dots,x_n) \in \mathbb{R}^n. \tag{2}$$

Here, $\ln_+(t) := \max(0, \ln t)$ $(t \ge 0)$. One can verify that the space $(\mathbb{R}^n, \|\cdot\|_{\log,n})$ is isomorphic (with an absolute constant) to the *dual* space for $(\mathbb{R}^n, \|\cdot\|_{\psi,n})$. More properties of these norms will be considered later. Now, let us state the spectral gap theorem for directed graphs in full generality:

Theorem C. *For every $\alpha \in (0,1)$, $m \in \mathbb{N}$ and $K > 0$ there are $L = L(\alpha,m,K) > 0$ and $n_0 = n_0(\alpha,m,K) \in \mathbb{N}$ with the following property: Let $n \ge n_0$, and let $\mathbf{d}^{in}, \mathbf{d}^{out}$ be two degree sequences such that for some integer $n^\alpha \le d \le 0.501n$ we have*

$$\max\big(\big\|\big(\mathbf{d}_i^{in} - d\big)_{i=1}^n\big\|_{\psi,n}, \big\|\big(\mathbf{d}_i^{out} - d\big)_{i=1}^n\big\|_{\psi,n}\big) \le K\sqrt{d}.$$

Assume that $\mathcal{D}_n(\mathbf{d}^{in}, \mathbf{d}^{out})$ is non-empty, let $\mathbf{G}$ be a random directed graph uniformly distributed on $\mathcal{D}_n(\mathbf{d}^{in}, \mathbf{d}^{out})$. Then

$$s_2(\mathbf{G}) \le L\sqrt{d}$$

with probability at least $1 - n^{-m}$.

The condition on the degree sequences in the theorem can be viewed as a concentration inequality for $\mathbf{d}_i^{in} - d$ and $\mathbf{d}_i^{out} - d$, with respect to the "uniform" choice of i in $[n]$. In particular, if $\|(\mathbf{d}_i^{in} - d)_{i=1}^n\|_\infty, \|(\mathbf{d}_i^{out} - d)_{i=1}^n\|_\infty \le K\sqrt{d}$ then the degree sequences satisfy the assumptions of the theorem.

Theorem C is the main theorem in this paper, and Theorem A (and, of course, B) is obtained as its consequence. In note [26], we proved a rather general comparison theorem for *jointly exchangeable* matrices which, in particular, allows us to estimate the spectral gap of random undirected d-regular graphs in terms of the second singular value of directed random graphs with predefined degree sequences. Let us briefly describe the idea of the reduction scheme. Assume that $\mathbf{G}$ is uniformly distributed on $\mathcal{G}_n(d)$ and let $\mathbf{M}$ be its adjacency matrix. Then the results of [26] assert that with high probability $s_2(\mathbf{M}) = \max(|\lambda_2(\mathbf{G})|, |\lambda_n(\mathbf{G})|)$ can be bounded from above by a multiple of the second largest singular value of the $n/2 \times n/2$ submatrix of $\mathbf{M}$ located in its top right corner. In turn, it can be verified that the distribution of this submatrix is directly related to the distribution of the adjacency matrix of a random *directed* graph on $n/2$ vertices with in- and out-degree sequences "concentrated" around $d/2$. We will cover this procedure in more detail in Section 6 and show how Theorem A follows from Theorem C.

In the course of proving Theorem C, we obtain certain relations for random graphs with predefined degree sequences which may be of separate interest. The rest of the introduction is devoted to discussing these developments and, in parallel, provides an outline of the proof of Theorem C. Given an $n \times n$ matrix M, we denote the Hilbert–Schmidt norm of M by $\|M\|_{HS}$. Additionally, we will write $\|M\|_\infty$ for the maximum norm (defined as the absolute value of the largest matrix entry). The set of adjacency matrices of graphs in $\mathcal{D}_n(\mathbf{d}^{in}, \mathbf{d}^{out})$ will be denoted by $\mathcal{M}_n(\mathbf{d}^{in}, \mathbf{d}^{out})$. Obviously, $\mathcal{M}_n(\mathbf{d}^{in}, \mathbf{d}^{out})$ coincides with the set of all 0-1-matrices M with $|\mathrm{supp}\,\mathrm{col}_i(M)| = \mathbf{d}_i^{in}$, $|\mathrm{supp}\,\mathrm{row}_i(M)| = \mathbf{d}_i^{out}$ for all $i \le n$.

The proof of Theorem C is composed of two major blocks. In the first block, we derive a concentration inequality for linear functionals of the form $\sum_{i,j=1}^n \mathbf{M}_{ij} Q_{ij}$, where $\mathbf{M}$ is a random matrix uniformly distributed on $\mathcal{M}_n(\mathbf{d}^{in}, \mathbf{d}^{out})$, and Q is any fixed $n \times n$ matrix. In the second block, we use the concentration inequality to establish certain discrepancy properties of the random graph associated with $\mathbf{M}$. Then, we apply a well known argument of Kahn and Szemerédi [18] in which the discrepancy property, together with certain covering arguments, yields a bound on the matrix norm.

The first block. Our concentration inequality for linear forms involves conditioning on a special event, having a probability close to one, on the space of matrices $\mathcal{M}_n(\mathbf{d}^{in}, \mathbf{d}^{out})$. Let us momentarily postpone the definition of the event (which is rather technical) and state the inequality first. Define a function $H(t)$ on the positive semi-axis as

$$H(t) := (1+t)\ln(1+t) - t. \tag{3}$$

Theorem D. *For every $\alpha \in (0,1)$, $m \in \mathbb{N}$ and $K > 0$ there are $\gamma = \gamma(\alpha, m, K)$, $L = L(\alpha, m, K) > 0$ and $n_0 = n_0(\alpha, m, K) \in \mathbb{N}$ with the following property: Let $n \ge n_0$, and let $\mathbf{d}^{in}, \mathbf{d}^{out}$ be two degree sequences such that for some integer $n^\alpha \le d \le 0.501n$ we have*

$$\max\left(\left\|\left(\mathbf{d}_i^{in} - d\right)_{i=1}^n\right\|_{\psi,n},\ \left\|\left(\mathbf{d}_i^{out} - d\right)_{i=1}^n\right\|_{\psi,n}\right) \le K\sqrt{d}.$$

Assume further that $\mathcal{M}_n(\mathbf{d}^{in}, \mathbf{d}^{out})$ is non-empty, and let $\mathbf{M}$ be uniformly distributed on $\mathcal{M}_n(\mathbf{d}^{in}, \mathbf{d}^{out})$. Then for any fixed $n \times n$ matrix Q and any $t \ge CL\sqrt{d}\|Q\|_{HS}$ we have

$$\mathbb{P}\Big\{\Big|\sum_{i,j=1}^n \mathbf{M}_{ij} Q_{ij} - \frac{d}{n}\sum_{i,j=1}^n Q_{ij}\Big| > t \ \mid\ \mathbf{M} \in \mathcal{E}_{\mathcal{P}}(L)\Big\} \le 2\exp\left(-\frac{d\,\|Q\|_{HS}^2}{n\,\|Q\|_\infty^2}\, H\left(\frac{\gamma t n \|Q\|_\infty}{d\|Q\|_{HS}^2}\right)\right).$$

Here, $C > 0$ is a universal constant and $\mathcal{E}_{\mathcal{P}}(L)$ is a subset of $\mathcal{M}_n(\mathbf{d}^{in}, \mathbf{d}^{out})$ which is determined by the value of L, and satisfies $\mathbb{P}(\mathcal{E}_{\mathcal{P}}(L)) \ge 1 - n^{-m}$.

The function H in the above deviation bound is quite natural in this context. It implicitly appears in the classical inequality of Bennett for sums of independent variables (see, [5, Formula 8b]), and later in the well known paper of Freedman [15] where he extends Bennett's inequality to martingales. In fact, our proof of Theorem D uses the Freedman inequality (more precisely, Freedman's bound for the moment generating function) as a fundamental element. Note that we require t to be greater (by the order of magnitude) than $\sqrt{d}\,\|Q\|_{HS}$, which makes the above statement a large deviation inequality. The restriction on t takes its roots into the way we obtain Theorem D from concentration inequalities for *individual* matrix rows. The *tensorization* procedure involves estimating the differences between conditional and unconditional expectations of rows, and we apply a rather crude bound by summing up absolute values of the "errors" for individual rows. In fact, the lower bound $CL\sqrt{d}\,\|Q\|_{HS}$ for t can be replaced with a smaller quantity $C'L\sqrt{d}\sum_{i=1}^n \|\mathrm{row}_i(Q)\|_{\log,n}$, provided that we choose a different "point of concentration" than $\frac{d}{n}\sum_{i,j=1}^n Q_{ij}$; we prefer to avoid discussing these purely technical aspects in the introduction.

A concentration inequality very similar to the one from Theorem D, was established in a recent paper of Cook, Goldstein and Johnson [12] which strongly influenced our work. The Bennett-type inequality from [12], formulated for adjacency matrices of undirected d-regular graphs, also involves a restriction on the parameter t, which, however, exhibits a completely different behavior compared to the lower bound $\sqrt{d}\,\|Q\|_{HS}$ in our work. In particular, the concentration inequality in [12] is not strong enough in the range $d \gg n^{2/3}$ to yield the correct order of the second largest eigenvalue. For the *permutation model*, a Bernstein-type concentration inequality was obtained in [13] by constructing a single martingale sequence for the whole matrix and applying Freedman's inequality. We will discuss in detail in Section 4 why a direct use of the same approach is problematic in our setting.

Theorem D, the way it is stated, is already sufficient to complete the proof of Theorem C, without any knowledge of the structure of the event $\mathcal{E}_{\mathcal{P}}(L)$. However, defining this event explicitly should give more insight and enable us to draw a comprehensive picture. Let G be a digraph on n vertices with degree sequences $\mathbf{d}^{in}$, $\mathbf{d}^{out}$, and let $M = (M_{ij})$ be the adjacency matrix of G. Further, let I be a subset of $\{1, 2, \dots, n\}$ (possibly, empty). We define quantities $p_j^{col}(I, M)$ and $p_j^{row}(I, M)$ $(j \le n)$ as

$$p_j^{col}(I, M) := \mathbf{d}_j^{in} - |\{q \in I:\ M_{qj} = 1\}| = |\{q \in I^c:\ M_{qj} = 1\}|;$$
$$p_j^{row}(I, M) := \mathbf{d}_j^{out} - |\{q \in I:\ M_{jq} = 1\}| = |\{q \in I^c:\ M_{jq} = 1\}|.$$

Further, let us define n-dimensional vectors $\mathcal{P}^{col}(I, M) = (\mathcal{P}_1^{col}(I, M), \dots, \mathcal{P}_n^{col}(I, M))$ and $\mathcal{P}^{row}(I, M) = (\mathcal{P}_1^{row}(I, M), \dots, \mathcal{P}_n^{row}(I, M))$ as

$$\left.\begin{aligned} \mathcal{P}_j^{col}(I, M) &:= \sum_{\ell=1}^n |p_j^{col}(I, M) - p_\ell^{col}(I, M)| \\ \mathcal{P}_j^{row}(I, M) &:= \sum_{\ell=1}^n |p_j^{row}(I, M) - p_\ell^{row}(I, M)| \end{aligned}\right. \qquad j \le n.$$

Conceptually, the vectors $\mathcal{P}^{col}(I, M), \mathcal{P}^{row}(I, M)$ can be thought of as a measure of "disproportion" in the locations of 1's across the matrix M. Given any non-empty subset $I \subset \{1, 2, \dots, n\}$, let M^{I^c} be the $I^c \times n$-submatrix of M. Then for every $j \le n$, $\mathcal{P}_j^{col}(I, M)$ is just the sum of differences of ℓ_1^n-norms of the j-th column and every other column of M^{I^c}:

$$\mathcal{P}_j^{col}(I, M) = \sum_{\ell=1}^{n} \big| \|\mathrm{col}_j(M^{I^c})\|_1 - \|\mathrm{col}_\ell(M^{I^c})\|_1 \big|.$$

The event $\mathcal{E}_{\mathcal{P}}(L)$ employed in Theorem D, controls the magnitude of those vectors: for every $L > 0$ we define the event as

$$\mathcal{E}_{\mathcal{P}}(L) := \Big\{ M \in \mathcal{M}_n(\mathbf{d}^{in}, \mathbf{d}^{out}) : \ \|\mathcal{P}^{row}(I, M)\|_{\psi,n}, \|\mathcal{P}^{col}(I, M)\|_{\psi,n} \le Ln\sqrt{d} \ \text{ for any interval subset } I \subset \{1, 2, \dots, n\} \text{ of cardinality at most } 0.001n \Big\}, \tag{4}$$

where $\| \cdot \|_{\psi,n}$ is given by (1). Note that the subsets I in the definition are assumed to be *interval subsets*, which gives importance to the way we enumerate the vertices. It is not difficult to see that if the definition involved *every* subset I with $|I| \le 0.001n$ then the probability of the event would be just zero as one can always find two vertices with largely non-overlapping sets of in-neighbors.

Loosely speaking, the condition secured by the event $\mathcal{E}_{\mathcal{P}}(L)$ is a skeleton for our matrix: it indicates that 1's are spread throughout the matrix more or less evenly. Assuming this property (i.e. conditioning on the event), we can establish stronger “rules” for the distribution of the non-zero elements and, in particular, obtain Theorem D.

From the technical perspective, the proof of Theorem D requires many preparatory statements and is quite long. Our exposition is largely self-contained; probably the only essential “exterior” statement which we employ in the first part of the paper is Freedman's inequality for martingales, which is given (together with some corollaries) in Sub-section 2.2. It is followed by the “graph” Subsection 2.3 where we state and prove a rough bound on the number of common in-neighbors of two vertices of a random graph using a standard argument involving simple switchings and multimaps (relations). Section 3 is the core of the paper. There, we apply the Freedman inequality and derive deviation bounds for individual rows of our random adjacency matrix. The first sub-section contains a series of lemmas dealing with a fixed row coordinate (and conditioned on the upper rows and all previous coordinates within this fixed row) and provides a foundation for our analysis. Sub-section 3.2 integrates the information for the individual matrix entries and, after resolving some technical issues, culminates in Theorem 3.12 which is the main statement of Section 3. Finally, we apply a tensorization procedure in Section 4 and prove (a somewhat technical version of) Theorem D.

The second block. Equipped with the concentration inequality given by Theorem D, we follow the Kahn–Szemerédi argument [18] to prove Theorem C. For simplicity, let us describe the procedure for the uniform model on $\mathcal{D}_n(d)$ and disregard conditioning on the event $\mathcal{E}_{\mathcal{P}}(L)$ in Theorem D. Denoting by $\mathbf{M}$ the adjacency matrix of a random d-regular graph uniformly distributed on $\mathcal{D}_n(d)$, it is easy to see that its largest singular value is equal d (deterministically), and the corresponding normalized singular vector is $(1/\sqrt{n}, 1/\sqrt{n}, \dots, 1/\sqrt{n}) = \frac{1}{\sqrt{n}}\mathbf{1}$. By the Courant–Fischer formula and the singular value decomposition, we have

$$s_2(\mathbf{M}) = \big\|\mathbf{M} - \frac{d}{n}\,\mathbf{1}\cdot\mathbf{1}^t\big\|_{2\to 2} = \sup_{\substack{x\in\mathbf{1}^\perp\cap S^{n-1},\\ y\in S^{n-1}}} \langle \mathbf{M}x, y\rangle.$$

A natural approach to bounding the supremum on the right hand side would be to apply the standard covering argument, which plays a key role in Asymptotic Geometric Analysis. The argument consists in showing first that $\langle \mathbf{M}x, y\rangle$ is bounded by certain threshold value (in this case, $O(\sqrt{d})$) with high probability for any pair of admissible x, y. Once this is done, a quite general approximation scheme allows to replace the supremum over $\mathbf{1}^\perp\cap S^{n-1}\times S^{n-1}$ by the supremum over a finite discrete subset (a net). From the probabilistic viewpoint, we pay the price by taking the union bound over the net (which can be chosen to have cardinality exponential in dimension) to obtain an estimate for the entire set. In order for such a procedure to work, we need a concentration inequality for $\langle \mathbf{M}x, y\rangle$ (for fixed x, y) which would "survive" multiplication by the cardinality of the net. By Theorem D (applied to the matrix $Q = yx^t$ for any fixed $(x, y) \in \mathbf{1}^\perp\cap S^{n-1}\times S^{n-1}$), we have

$$\mathbb{P}\big\{|\langle \mathbf{M}x, y\rangle| \gg \sqrt{d}\big\} \ll \exp\left(-\frac{d}{n\,\|x\|_\infty^2\,\|y\|_\infty^2}\, H\left(\frac{n\,\|x\|_\infty\,\|y\|_\infty}{\sqrt{d}}\right)\right).$$

However, the expression on the right hand side is an increasing function of $\|x\|_\infty\,\|y\|_\infty$, and becomes larger than C^{-n} when $\|x\|_\infty\,\|y\|_\infty \gg \sqrt{d}/n$. Hence, the union bound in the above description can work only for x, y having small $\|\cdot\|_\infty$-norms. A key idea in the argument by Kahn and Szemerédi, which distinguishes it from the standard covering procedure, is to split the quadratic form associated with $\langle \mathbf{M}x, y\rangle$ into "flat" and "spiky" parts:

$$\langle \mathbf{M}x, y\rangle = \sum_{\substack{(i,j)\in[n]\times[n]:\\ |x_j y_i|\le\sqrt{d}/n}} y_i\mathbf{M}_{ij}x_j + \sum_{\substack{(i,j)\in[n]\times[n]:\\ |x_j y_i|>\sqrt{d}/n}} y_i\mathbf{M}_{ij}x_j. \tag{5}$$

Let us note that a somewhat similar decomposition of the sphere into "flat" and "spiky" vectors was used in [20] and [24] to bound the smallest singular value of certain random matrices. The first term in (5) can be dealt with by directly using the concentration inequality from Theorem D (plus standard covering). On the other hand, the second summand needs a more delicate handling. Kahn and Szemerédi proposed a way to relate the quantity to discrepancy properties of the underlying graph, more precisely, to deviations of the edge count between subsets of the vertices from its mean value. To illustrate the connection, let a, b be any positive numbers with $ab \gg \sqrt{d}/n$ and let $S := \{i \le n : |y_i| \approx b\}$ and

$T := \{j \leq n : |x_j| \approx a\}$. Then

$$\sum_{\substack{(i,j)\in[n]\times[n]: \\ |x_j|\approx a, |y_i|\approx b}} y_i \mathbf{M}_{ij} x_j = O\big(ab\,|\mathbf{E}_G(S,T)|\big),$$

where $|\mathbf{E}_G(S,T)|$ is the number of edges of graph $\mathbf{G}$ corresponding to $\mathbf{M}$, starting in S and ending in T. In the actual proof, this simplified illustration should be replaced by a careful partitioning of vectors x and y into "almost constant" blocks. We refer to Section 5 for a rigorous exposition of the argument allowing to complete the proof of Theorem C. Once Theorem C is proved, we apply it, together with the "de-symmetrization" result of [26], to prove Theorem A. This is accomplished in Section 6.

2. Notation and Preliminaries

Everywhere in the text, we assume that n is a large enough natural number. For a finite set I, by $|I|$ we denote its cardinality. For any positive integer m, the set $\{1, 2, \dots, m\}$ will be denoted by $[m]$. If $I \subset [n]$ then, unless explicitly specified otherwise, the set I^c is the complement of I in $[n]$. For a real number a, $\lceil a \rceil$ is the smallest integer greater or equal to a, and $\lfloor a \rfloor$ is the largest integer not exceeding a. A vector $y \in \mathbb{R}^n$ is called *r-sparse* for some $r \geq 0$ if the support $\operatorname{supp} y$ has cardinality at most r. By $\langle \cdot, \cdot \rangle$ we denote the standard inner product in $\mathbb{R}^n$, by $\|\cdot\|$ — the standard Euclidean norm in $\mathbb{R}^n$, and by $\{e_1, e_2, \dots, e_n\}$ — the canonical basis vectors. For every $1 \leq p < \infty$, the $\|\cdot\|_p$-norm in $\mathbb{R}^n$ is defined by

$$\|(x_1, x_2, \dots, x_n)\|_p := \left(\sum_{i=1}^{\infty} |x_i|^p\right)^{1/p},$$

and the canonical maximal norm is

$$\|(x_1, x_2, \dots, x_n)\|_\infty := \max_{i \leq n} |x_i|.$$

Universal constants are denoted by C, c, c', etc. In some situations we will add a numerical subscript to the name of a constant to relate it to a particular numbered statement. For example, $C_{2.2}$ is a constant from Lemma 2.2.

Let M be a fixed $n \times n$ matrix. The (i,j)-th entry of M is denoted by M_{ij}. Further, we will denote rows and columns by $\operatorname{row}_1(M), \dots, \operatorname{row}_n(M)$ and $\operatorname{col}_1(M), \dots, \operatorname{col}_n(M)$. We denote the Hilbert–Schmidt norm of M by $\|M\|_{HS}$. Additionally, we write $\|M\|_\infty$ for the maximum norm (defined as the absolute value of the largest matrix entry) and $\|M\|_{2\to 2}$ for its spectral norm.

Let $\mathbf{d}^{in}, \mathbf{d}^{out}$ be two degree sequences. Everywhere in this paper, we assume that for an integer d we have

$$(1-c_0)d \leq \mathbf{d}^{in}_i, \mathbf{d}^{out}_i \leq d \quad \text{for all } i \leq n, \quad \text{where } c_0 := 0.001 \text{ and } d \leq (1/2+c_0)n. \tag{6}$$

Recall that, given two degree sequences $\mathbf{d}^{in}, \mathbf{d}^{out}$, the set of adjacency matrices of graphs in $\mathcal{D}_n(\mathbf{d}^{in}, \mathbf{d}^{out})$ is denoted by $\mathcal{M}_n(\mathbf{d}^{in}, \mathbf{d}^{out})$. We will write $\mathcal{S}_n(d)$ for the set of adjacency matrices of undirected simple d-regular graphs on $[n]$. Each of the sets $\mathcal{D}_n(\mathbf{d}^{in}, \mathbf{d}^{out})$, $\mathcal{M}_n(\mathbf{d}^{in}, \mathbf{d}^{out})$, $\mathcal{G}_n(d)$, $\mathcal{S}_n(d)$ can be turned into a probability space by defining the normalized counting measure. We will use the same notation $\mathbb{P}$ for the measure in each of the four cases. The actual probability space will always be clear from the context.

The expectation of a random variable ξ is denoted by $\mathbb{E}\xi$. We will use vertical bar notation for conditional expectation and conditional probability. For example, the expectation of ξ conditioned on an event $\mathcal{E}$, will be written as $\mathbb{E}[\xi \,|\, \mathcal{E}]$, and the conditional expectation given a σ-sub-algebra $\mathcal{F}$ — as $\mathbb{E}[\xi \,|\, \mathcal{F}]$.

Let A, B be sets, and $R \subset A \times B$ be a relation. Given $a \in A$ and $b \in B$, the image of a and preimage of b are defined by

$$R(a) := \{y \in B \,:\, (a, y) \in R\} \quad \text{and} \quad R^{-1}(b) := \{x \in A \,:\, (x, b) \in R\}.$$

We also set $R(A) := \cup_{a \in A} R(a)$. Further in the text, we will define relations between sets in order to estimate their cardinality, using the following elementary claim (see [19] for a proof):

Claim 2.1. *Let $s, t > 0$. Let R be a relation between two finite sets A and B such that for every $a \in A$ and every $b \in B$ one has $|R(a)| \geq s$ and $|R^{-1}(b)| \leq t$. Then $s|A| \leq t|B|$.*

2.1. Orlicz norms

In the Introduction, we defined two Orlicz norms $\|\cdot\|_{\psi,n}$ and $\|\cdot\|_{\log,n}$ in $\mathbb{R}^n$. Let us state some of their elementary properties (see [23] for extensive information on Orlicz functions and Orlicz spaces). First, it can be easily checked that

$$\|x\|_{\psi,n} \leq \|x\|_\infty \leq \ln(en)\, \|x\|_{\psi,n} \qquad \text{for all } x \in \mathbb{R}^n. \tag{7}$$

Similarly, we have

$$\frac{n}{\ln n}\|x\|_{\log,n} \leq \|x\|_1 \leq en\|x\|_{\log,n} \qquad \text{for all } x \in \mathbb{R}^n. \tag{8}$$

Lemma 2.2. *For any vector $y \in \mathbb{R}^n$ with $m := |\operatorname{supp} y| \leq n$ we have*

$$\|y\| \leq C_{2.2}\sqrt{m}\|y\|_{\psi,n} \ln \frac{2n}{m},$$

where $C_{2.2} > 0$ is a universal constant.

Proof. Without loss of generality, $z := \|y\|_{\psi,n} = \sqrt{n/m}$. The convex conjugate of the exponential function is $t\ln(t) - t$ ($t > 0$). Hence, by Fenchel's inequality, for any $i \in \operatorname{supp} y$ we have

$$\begin{aligned} {y_i}^2 &\le e^{|y_i|/z} + z|y_i|\ln(z|y_i|) - z|y_i| \\ &\le e^{|y_i|/z} + z|y_i|\ln(2z^2) + z|y_i|\ln\left(\frac{|y_i|}{2z}\right) \\ &\le e^{|y_i|/z} + z|y_i|\ln(2z^2) + \frac{1}{2}{y_i}^2. \end{aligned}$$

Summing over all $i \in \operatorname{supp} y$, we get

$$\|y\|^2 \le 2en + 2z\ln(2z^2)\sum_{i\in\operatorname{supp} y}|y_i| \le 2en + 2z\ln(2z^2)\sqrt{m}\|y\|.$$

Plugging in the definition of z and solving the above inequality, we get

$$\|y\| \le C\sqrt{n}\ln\frac{2n}{m}$$

for some universal constant $C > 0$. The result follows. □

By a duality argument, we also have the following.

Lemma 2.3. *For any vector $y \in \mathbb{R}^n$ with $m := |\operatorname{supp} y| \le n$ we have*

$$n\|y\|_{\log,n} \le C_{2.3}\|y\|\sqrt{m}\ln\frac{2n}{m} \le C_{2.3}\|y\|_1\ln\frac{2n}{m},$$

where $C_{2.3} > 0$ is a universal constant.

Finally, given a vector x with $\|x\|_{\psi,n} = 1$, we can bound the number of coordinates x of any given magnitude:

Lemma 2.4. *Let $x \in \mathbb{R}^n$ with $\|x\|_{\psi,n} = 1$. Then there is a natural number $k \le 2\ln(en)$ such that*

$$\left|\left\{i \le n : |x_i| \ge k/2\right\}\right| \ge n(2e)^{-k}.$$

Proof. By the definition of the norm $\|\cdot\|_{\psi,n}$, we have

$$\sum_{i=1}^n e^{|x_i|} = en,$$

whence

$$\sum_{\substack{i\le n:\\ |x_i|\ge 1/2}} e^{|x_i|} \ge n.$$

Thus,

$$\sum_{k=1}^\infty \left|\left\{i \le n : |x_i| \ge k/2\right\}\right| e^{(k+1)/2} \ge \sum_{k=1}^\infty 2^{-k}n.$$

It remains to note that, in view of (7), we have $\left\{i \le n : |x_i| \ge k/2\right\} = \emptyset$ for all $k > 2\ln(en)$ and that $2^{-k}e^{-(k+1)/2} \ge (2e)^{-k}$ for all k. □

2.2. Freedman's inequality

In this sub-section, we recall the classical concentration inequality for martingales due to Freedman, and provide several auxiliary statements which we will apply later in Section 4. Define

$$g(t) := e^t - t - 1, \quad t > 0. \tag{9}$$

In [15], Freedman proved the following bound for the moment-generating function which will serve as a fundamental block of this paper:

Theorem 2.5 (Freedman's inequality)**.** *Let $m \in \mathbb{N}$, let $(X_i)_{i\le m}$ be a martingale with respect to a filtration $(\mathcal{F}_i)_{i\le m}$, and let*

$$d_i := X_i - X_{i-1}, \;\; i \le m,$$

be the corresponding difference sequence. Assume that $|d_i| \le M$ a.s. for some $M > 0$ and $\sum_{i=1}^m \mathbb{E}({d_i}^2 \,|\, \mathcal{F}_{i-1}) \le \sigma^2$ a.s. for some $\sigma > 0$. Then for any $\lambda > 0$, we have

$$\mathbb{E}e^{\lambda(X_m - X_0)} \le \exp\left(\frac{\sigma^2}{M^2}\, g(\lambda M)\right).$$

As a consequence of the above relation, Freedman derived the inequality

$$\mathbb{P}\big\{X_m - X_0 \ge t\big\} \le \exp\left(-\frac{\sigma^2}{M^2} H\Big(\frac{Mt}{\sigma^2}\Big)\right), \quad t > 0, \tag{10}$$

where H is defined by (3). It is easy to check that

$$H(t) \ge \frac{t^2}{2(1+t/3)} \quad \text{for any } t \ge 0, \tag{11}$$

whence, with the above notation,

$$\mathbb{P}\big\{X_m - X_0 \ge t\big\} \le \exp\left(-\frac{t^2}{2\sigma^2 + 2Mt/3}\right), \quad t > 0. \tag{12}$$

In the special case when the martingale consists of partial sums of a series of i.i.d. centered random variables, i.e. d_i ($i \le m$) are i.i.d., (10) was obtained by Bennett [5] and (12) derived by Bernstein [6]. Returning to arbitrary martingale sequences, the estimate (10) is often referred to as *the Freedman inequality*. However, in our setting it is crucial to have the stronger relation provided by Theorem 2.5, as it will allow us to *tensorize* concentration inequalities obtained for individual rows of the matrix.

Lemma 2.6. *Let $m \in \mathbb{N}$ and let $\xi_1, \xi_2, \dots, \xi_m$ be random variables. Further, assume that $f_i(\lambda) : \mathbb{R}_+ \to \mathbb{R}_+$ are functions such that*

$$\mathbb{E}[e^{\lambda \xi_i} \mid \xi_1, \dots, \xi_{i-1}] \le f_i(\lambda)$$

for any $\lambda > 0$ and $i \le m$. Then for any subset $T \subset [m]$ we have

$$\mathbb{E}e^{\lambda \sum_{i\in T} \xi_i} \le \prod_{i\in T} f_i(\lambda).$$

Proof. Without loss of generality, take $T = [m]$. Note that

$$\mathbb{E}\, e^{\lambda \sum_{i=1}^m \xi_i} = \mathbb{E}\left[\mathbb{E}[e^{\lambda \sum_{i=1}^m \xi_i} \mid \xi_1, \dots, \xi_{m-1}]\right] = \mathbb{E}\left[e^{\lambda \sum_{i=1}^{m-1} \xi_i}\, \mathbb{E}[e^{\lambda \xi_m} \mid \xi_1, \dots, \xi_{m-1}]\right].$$

Hence, by the assumption on f_m, we get

$$\mathbb{E}\, e^{\lambda \sum_{i=1}^m \xi_i} \le f_m(\lambda)\, \mathbb{E}\, e^{\lambda \sum_{i=1}^{m-1} \xi_i}.$$

Iterating this procedure, we obtain

$$\mathbb{E}\, e^{\lambda \sum_{i=1}^m \xi_i} \le \prod_{i=1}^m f_i(\lambda).$$

□

As a corollary, we obtain a tail estimate for the sum of random variables satisfying a "Freedman type" bound for their moment generating functions.

Corollary 2.7. *Let $m \in \mathbb{N}$; let $(M_i)_{i \le m}$ and $(\sigma_i)_{i \le m}$ be two sequences of positive numbers and let random variables $\xi_1, \dots, \xi_m$ satisfy*

$$\mathbb{E}[e^{\lambda \xi_i} \mid \xi_1, \dots, \xi_{i-1}] \le \exp\left(\frac{{\sigma_i}^2}{{M_i}^2}\, g(\lambda M_i)\right).$$

for any $i \le m$ and $\lambda \ge 0$. Then for any $t \ge 0$, we have

$$\mathbb{P}\Big\{\sum_{i \le m} \xi_i \ge t\Big\} \le \exp\left(-\frac{\sigma^2}{M^2}\, H\left(\frac{tM}{\sigma^2}\right)\right),$$

where $M := \max_{i \le m} M_i$ and $\sigma^2 := \sum_{i=1}^m {\sigma_i}^2$.

Proof. Fix any $t > 0$ and set $\lambda := \ln(1 + tM/\sigma^2)/M$. In view of the assumptions on ξ_i's and Lemma 2.6, we have

$$\mathbb{E}\, e^{\lambda \sum_{i=1}^m \xi_i} \le \prod_{i=1}^m \exp\left(\frac{{\sigma_i}^2}{{M_i}^2}\, g(\lambda M_i)\right).$$

Since the function $g(\lambda t)/t^2$ is increasing on $(0, \infty)$, the last relation implies

$$\mathbb{E}\, e^{\lambda \sum_{i=1}^m \xi_i} \le \prod_{i=1}^m \exp\left(\frac{{\sigma_i}^2}{M^2}\, g(\lambda M)\right) = \exp\left(\frac{\sigma^2}{M^2}\, g(\lambda M)\right).$$

Hence, by Markov's inequality,

$$\mathbb{P}\Big\{\sum_{i \le m} \xi_i \ge t\Big\} \le e^{-\lambda t}\, \mathbb{E}\, e^{\lambda \sum_{i=1}^m \xi_i} \le e^{-\lambda t}\, \exp\left(\frac{\sigma^2}{M^2}\, g(\lambda M)\right).$$

The result follows after plugging in the expression for λ. □

2.3. A crude bound on the number of common in-neighbors

We start this sub-section with some graph notations. Let $G = ([n], E)$ be a directed graph on $[n]$ with the edge set E and adjacency matrix M. For any vertex $i \in [n]$, we define the set of its in-neighbors

$$\mathcal{N}_G^{in}(i) := \big\{ v \le n : \; (v, i) \in E \big\} = \operatorname{supp} \operatorname{col}_i(M).$$

Similarly, the set of out-neighbors is

$$\mathcal{N}_G^{out}(i) := \big\{ v \le n : \, (i, v) \in E \big\} = \operatorname{supp} \operatorname{row}_i(M).$$

Further, for every $I, J \subset [n]$ the set of all edges departing from I and landing in J is denoted by

$$\mathbf{E}_G(I, J) := \big\{ e \in E : \, e = (i, j) \text{ for some } i \in I \text{ and } j \in J \big\}.$$

The set of common in-neighbors of two vertices u, v is

$$\mathbf{C}_{\mathrm{G}}^{\mathrm{in}}(u, v) := \{ i \le n \, : \, (i, u), (i, v) \in E \} = \operatorname{supp} \operatorname{col}_u(M) \cap \operatorname{supp} \operatorname{col}_v(M).$$

In this sub-section, we estimate the probability that a pair of distinct vertices of a random graph uniformly distributed on $\mathcal{D}_n(\mathbf{d}^{in}, \mathbf{d}^{out})$, has many common in-neighbors, conditioned on a special σ-algebra. Let us note that (much stronger) results of this type for *d-regular* directed graphs, as well as bipartite regular undirected graphs, were obtained in [10]. Unlike in [10], we are only interested in large deviations for $\mathbf{C}_{\mathrm{G}}^{\mathrm{in}}(i, j)$. On the other hand, the specifics of our setting is that our graphs are not regular (instead, have predefined in- and out-degree sequences) and that the probability is conditional. More precisely, given a subset $S \subset [n]$, let $\mathcal{F}$ be the σ-algebra on $\mathcal{D}_n(\mathbf{d}^{in}, \mathbf{d}^{out})$ with atoms of the form $\{G \in \mathcal{D}_n(\mathbf{d}^{in}, \mathbf{d}^{out}) : \, \mathbf{E}_G(S, [n]) = F\}$ for all subsets $F \subset [n] \times [n]$. In other words, each atom of $\mathcal{F}$ is a set of graphs sharing the same collection of out-edges for vertices in S. Then for any event $\mathcal{E} \subset \mathcal{D}_n(\mathbf{d}^{in}, \mathbf{d}^{out})$, we let $\mathbb{P}\big\{G \in \mathcal{E} \,|\, \mathbf{E}_G(S, [n])\big\}$ be the conditional probability of $\mathcal{E}$ given $\mathcal{F}$.

Let us remark that the proof of the main statement of this sub-section is a rather standard application of the method of *simple switchings* introduced by Senior [25] and developed by McKay and Wormald (see [21] as well as survey [30]). We provide the proof for the reader's convenience.

Proposition 2.8. *There exist universal constants $c_{2.8}, C_{2.8} > 0$ with the following property. Asssume that $C_{2.8} \ln n \le d \le (1/2 + c_0)n$, and let two degree sequences $\mathbf{d}^{in}, \mathbf{d}^{out} \in \mathbb{R}^n$ satisfy (6). Let $I \subset [n]$ be such that $|I| \le c_0 n$. Then, denoting by $\mathcal{E}_{2.8}$ the event*

$$\mathcal{E}_{2.8} := \Big\{ G \in \mathcal{D}_n(\mathbf{d}^{in}, \mathbf{d}^{out}) : \, \exists i \neq j, \, |\mathbf{C}_{\mathrm{G}}^{\mathrm{in}}(i, j) \cap I^c| \ge 0.9d \Big\},$$

we have

$$\mathbb{P}\big\{ G \in \mathcal{E}_{2.8} \,|\, \mathbf{E}_G(I, [n]) \big\} \le \exp(-c_{2.8} d).$$

For the rest of the sub-section, we will assume that d and I satisfy the assumptions of Proposition 2.8, and we restrict ourselves to an atom of the σ-algebra generated by $\mathbf{E}_G(I,[n])$. Namely, let $F \subset [n]\times[n]$ be such that the set of graphs from $\mathcal{D}_n(\mathbf{d}^{in},\mathbf{d}^{out})$ satisfying $\mathbf{E}_G(I,[n]) = F$, is non-empty. Given $1 \le i \neq j \le n$, we let

$$\mathcal{E}_{i,j} := \Big\{G \in \mathcal{D}_n(\mathbf{d}^{in},\mathbf{d}^{out}) :\, \mathbf{E}_G(I,[n]) = F,\, |\mathbf{C}^{\rm in}_{\rm G}(i,j) \cap I^c| \ge 0.9d\Big\}$$

and for any natural $q \ge 0.8d$, let

$$\mathcal{E}^q_{i,j} := \Big\{G \in \mathcal{D}_n(\mathbf{d}^{in},\mathbf{d}^{out}) :\, \mathbf{E}_G(I,[n]) = F,\, |\mathbf{C}^{\rm in}_{\rm G}(i,j) \cap I^c| = q\Big\}.$$

Lemma 2.9. *Let $G \in \mathcal{E}^q_{1,2}$ (for some $q \ge 0.8d$), $q' < q$, and denote*

$$J := \{j \ge 3 :\, |\mathbf{C}^{\rm in}_{\rm G}(1,2) \cap I^c \setminus \mathcal{N}^{in}_G(j)| \ge q'\}.$$

Let $\Phi_{1,2} := I^c \setminus \big(\mathcal{N}^{in}_G(1) \cup \mathcal{N}^{in}_G(2)\big)$. Then

$$|\mathbf{E}_G(\Phi_{1,2},J)| \ge dq\bigg(2 - c_0 - \frac{6c_0 d}{q} - \frac{d}{q-q'}\bigg).$$

Proof. First note that since $d \le (1/2 + c_0)n$ and the degree sequences satisfy (6), we have

$$|\Phi_{1,2}| \ge |I^c| - |\mathcal{N}^{in}_G(1) \cup \mathcal{N}^{in}_G(2)| \ge (1-c_0)n - 2d + q \ge q - 6c_0 d.$$

Therefore

$$|\mathbf{E}_G(\Phi_{1,2},[n])| \ge |\Phi_{1,2}| \min_{i\in\Phi_{1,2}} \mathbf{d}^{out}_i \ge (1-c_0)d(q-6c_0d). \tag{13}$$

In view of the definition of J, for any $j \in J^c$ we have

$$|\mathbf{C}^{\rm in}_{\rm G}(1,2) \cap I^c \cap \mathcal{N}^{in}_G(j)| \ge q - q'.$$

Hence,

$$(q-q')|J^c| \le |\mathbf{E}_G(\mathbf{C}^{\rm in}_{\rm G}(1,2)\cap I^c, J^c)| \le |\mathbf{E}_G(\mathbf{C}^{\rm in}_{\rm G}(1,2)\cap I^c, [n])| \le qd,$$

which implies that $|J^c| \le qd/(q-q')$. On the other hand, for every $j \in J^c$ we have

$$|\Phi_{1,2} \cap \mathcal{N}^{in}_G(j)| \le |\mathcal{N}^{in}_G(j)| - |\mathbf{C}^{\rm in}_{\rm G}(1,2) \cap I^c \cap \mathcal{N}^{in}_G(j)| \le d - q + q',$$

whence

$$|\mathbf{E}_G(\Phi_{1,2},J^c)| \le |J^c|(d-q+q') \le qd\bigg(\frac{d}{q-q'} - 1\bigg).$$

Together with (13), this gives the result. □

Lemma 2.10. *For any integer $q \geq 0.8d+1$, we have $|\mathcal{E}_{1,2}^q| \leq 0.9\,|\mathcal{E}_{1,2}^{q-1}|$.*

Proof. Let us define a relation R on $\mathcal{E}_{1,2}^q \times \mathcal{E}_{1,2}^{q-1}$ as follows:

Pick any $G \in \mathcal{E}_{1,2}^q$, and choose an edge $(i,j) \in \mathbf{E}_G(\Phi_{1,2}, J)$ and $k \in \mathbf{C}_{\mathrm{G}}^{\mathrm{in}}(1,2) \cap I^c \setminus \mathcal{N}_G^{in}(j)$, where J and $\Phi_{1,2}$ are defined in Lemma 2.9 with $q' := \lceil q/7 \rceil$. Perform the simple switching on the graph G, replacing the edges (i,j) and $(k,1)$ with $(i,1)$ and (k,j) respectively. Note that the conditions $i \notin \mathcal{N}_G^{in}(1)$ and $k \notin \mathcal{N}_G^{in}(j)$ guarantee that the simple switching does not create multiple edges. Moreover, since $i \in \Phi_{1,2}$, we obtain a valid graph $G' \in \mathcal{E}_{1,2}^{q-1}$. We define $R(G)$ as the set of all graphs G' which can be obtained from G via the above procedure.

Using Lemma 2.9 and the definition of J, we get

$$|R(G)| \geq |\mathbf{E}_G(\Phi_{1,2}, J)| \cdot \min_{j \in J} |\mathbf{C}_{\mathrm{G}}^{\mathrm{in}}(1,2) \cap I^c \setminus \mathcal{N}_G^{in}(j)| \geq \frac{1}{7} dq^2 \left(2 - c_0 - \frac{6c_0 d}{q} - \frac{d}{q - \lceil q/7 \rceil}\right). \tag{14}$$

Now we estimate the cardinalities of preimages. Let $G' \in R(\mathcal{E}_{1,2}^q)$. In order to reconstruct a graph G for which $(G, G') \in R$, we need to perform a simple switching which destroys an edge in $\mathbf{E}_{G'}(\mathcal{N}_{G'}^{in}(1) \cap I^c \setminus \mathbf{C}_{\mathrm{G'}}^{\mathrm{in}}(1,2), \{1\})$ and adds an edge connecting a vertex in $\mathcal{N}_{G'}^{in}(2) \cap I^c \setminus \mathbf{C}_{\mathrm{G'}}^{\mathrm{in}}(1,2)$ to vertex 1. There are at most $(\mathbf{d}_1^{in} - q + 1)$ choices to destroy an edge in $\mathbf{E}_{G'}(\mathcal{N}_{G'}^{in}(1) \cap I^c \setminus \mathbf{C}_{\mathrm{G'}}^{\mathrm{in}}(1,2), \{1\})$, and at most $(\mathbf{d}_2^{in} - q + 1)$ possibilities to add an edge connecting $\mathcal{N}_{G'}^{in}(2) \cap I^c \setminus \mathbf{C}_{\mathrm{G'}}^{\mathrm{in}}(1,2)$ to 1. Finally, there are at most d possibilities to complete the switching. Thus, by the assumptions on the degree sequences,

$$|R^{-1}(G')| \leq d(d-q+1)^2.$$

This, together with (14), the choice of q and the constant c_0, finishes the proof after using Claim 2.1. □

Proof of Proposition 2.8. Iterating the last lemma, we deduce that for any $q \geq 0.8d+1$ we have

$$|\mathcal{E}_{1,2}^q| \leq 0.9^{q - \lceil 0.8d \rceil}\, |\mathcal{E}_{1,2}^{\lceil 0.8d \rceil}| \leq 0.9^{q - \lceil 0.8d \rceil} \left|\left\{G \in \mathcal{D}_n(\mathbf{d}^{in}, \mathbf{d}^{out}) :\, \mathbf{E}_G(I, [n]) = F\right\}\right|.$$

Hence,

$$\mathbb{P}(\mathcal{E}_{1,2}) = \sum_{q \geq 0.9d} \mathbb{P}(\mathcal{E}_{1,2}^q) \leq 0.9^{0.1d-1}\, \mathbb{P}\left\{G \in \mathcal{D}_n(\mathbf{d}^{in}, \mathbf{d}^{out}) :\, \mathbf{E}_G(I, [n]) = F\right\}.$$

Similarly, we have

$$\mathbb{P}(\mathcal{E}_{i,j}) \leq 0.9^{0.1d-1}\, \mathbb{P}\left\{G \in \mathcal{D}_n(\mathbf{d}^{in}, \mathbf{d}^{out}) :\, \mathbf{E}_G(I, [n]) = F\right\}$$

for any $i \neq j$. Applying the union bound and the definition of $\mathcal{E}_{2.8}$, we deduce that

$$\mathbb{P}(\mathcal{E}_{2.8} \cap \{G :\, \mathbf{E}_G(I, [n]) = F\}) \leq n^2\, 0.9^{0.1d-1}\, \mathbb{P}\left\{G \in \mathcal{D}_n(\mathbf{d}^{in}, \mathbf{d}^{out}) :\, \mathbf{E}_G(I, [n]) = F\right\}.$$

The result follows in view of the assumptions on n and d. □

Remark 2.11. Let us emphasize that much sharper bounds on the number of common in-neighbors can be obtained by applying results proved later in this paper. However, not being the central subject of this work, no improvements to Proposition 2.8 will be pursued.

3. A concentration inequality for a matrix row

Take a large enough natural number n, two degree sequences $\mathbf{d}^{in}, \mathbf{d}^{out} \in \mathbb{R}^n$, satisfying (6), and a non-negative integer number $m \leq c_0 n$. Further, let $Y^1, Y^2, \dots, Y^m$ be $\{0,1\}$-vectors such that the set of matrices

$$\widetilde{\mathcal{M}_n} := \big\{M \in \mathcal{M}_n(\mathbf{d}^{in}, \mathbf{d}^{out}) : \mathrm{row}_i(M) = Y^i \text{ for all } i \leq m\big\}$$

is non-empty. The parameters n, $\mathbf{d}^{in}$, $\mathbf{d}^{out}$, m and $Y^1, Y^2, \dots, Y^m$ are fixed throughout this section. As we mentioned in Section 2, we always assume (6). Our goal here is to show that, under certain conditions on the degree sequences and vectors $Y^1, \dots, Y^m$, the $(m+1)$-st row of the random matrix uniformly distributed in the set $\widetilde{\mathcal{M}_n}$ enjoys strong concentration properties.

For each $\ell \leq n$, define

$$p_\ell := \mathbf{d}^{in}_\ell - |\{i \leq m : Y^i_\ell = 1\}|. \tag{15}$$

Everywhere in this section, we assume that vectors $Y^1, \dots, Y^m$ are such that p_ℓ's satisfy

$$\mathbf{d}^{in}_\ell \geq p_\ell \geq (1 - 2c_0)\mathbf{d}^{in}_\ell \quad \forall \ell \in [n]. \tag{16}$$

Note that the above condition implies $|p_\ell - p_{\ell'}| \leq |\mathbf{d}^{in}_\ell - \mathbf{d}^{in}_{\ell'}| + 4c_0 d$ for all $\ell, \ell' \in [n]$. Let us remark that in the second part of the section we will employ much stronger assumptions on p_ℓ.

Further, let Ω be the set of all $\{0,1\}$-vectors $v \in \mathbb{R}^n$ such that $|\mathrm{supp}\, v| = \mathbf{d}^{out}_{m+1}$. Then we can define an induced probability measure $\mathbb{P}_\Omega$ on Ω by setting

$$\mathbb{P}_\Omega(A) := \mathbb{P}\big\{\mathrm{row}_{m+1}(M) \in A \,|\, M \in \widetilde{\mathcal{M}_n}\big\}, \quad A \subset \Omega.$$

Let $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Consider the filtration of σ-algebras $\{\mathcal{F}_i\}_{i=0}^n$ on $(\Omega, \mathbb{P}_\Omega)$ which reveals the coordinates of the $(m+1)$-st row one by one, i.e. $\mathcal{F}_i$ is generated by $\mathcal{F}_{i-1}$ and by the variable v_i (where v is distributed on Ω according to the measure $\mathbb{P}_\Omega$) for any $i = 1, 2, \dots, n$.

3.1. Distribution of the i-th coordinate

Everywhere in this sub-section, we assume that the number d satisfies conditions of Proposition 2.8, i.e.

$$d \geq C_{2.8} \ln n.$$

We fix a number $i \leq n$ and numbers $\varepsilon_1, \varepsilon_2 \dots, \varepsilon_{i-1} \in \{0,1\}$ such that $\mathbb{P}_\Omega\{v \in \Omega : v_j = \varepsilon_j \text{ for all } j < i\} > 0$. Let us denote

$$Q := \{v \in \Omega : v_j = \varepsilon_j \text{ for all } j < i\}$$

and let $\mathbb{P}_Q := \mathbb{P}_\Omega(\cdot\,|\,Q)$ be the induced probability measure on Q. By $\mathcal{F}_k \cap Q$ we denote restrictions of the previously defined σ-algebras to Q. Obviously, $\mathcal{F}_k \cap Q = \{\emptyset, Q\}$ for all $k \leq i-1$.

The goal of the sub-section is to develop machinery for dealing with arbitrary *functions* on Q. Loosely speaking, given a function $h : Q \to \mathbb{R}$ satisfying certain conditions, we will study the "impact" of the i-th coordinate of its argument on its value. Then, in Sub-section 3.2, we will apply the relations established here, together with the Freedman inequality, to obtain concentration inequalities for the $(m+1)$-st row of a random matrix uniformly distributed on $\widetilde{\mathcal{M}_n}$. The central technical statement of this part of the paper is Lemma 3.8. On the way to stating and proving the lemma, we will go through several auxiliary statements and introduce several useful notions.

Lemma 3.1. *Let Q be as above, and let $k \neq \ell \in \{i, \dots, n\}$. Further, let vectors $v, v' \in Q$ be such that $\{j \leq n : v_j \neq v'_j\} = \{k, \ell\}$ with $v_k = v'_\ell = 1$ and $v_\ell = v'_k = 0$. Then*

$$\mathbb{P}_Q(v) \leq \gamma_{k,\ell}\, \mathbb{P}_Q(v'),$$

where

$$\gamma_{k,\ell} := \frac{1}{1 - e^{-c_{2.8} d}} \left[\frac{p_\ell}{p_k} \mathbf{1}_{p_\ell < p_k} + \Big(1 + \frac{p_\ell - p_k}{p_k - \lfloor 0.9d \rfloor}\Big) \mathbf{1}_{p_\ell \geq p_k} \right], \tag{17}$$

p_ℓ's are given by (15) and the constant $c_{2.8}$ is defined in Proposition 2.8.

Proof. First, let us define

$$\widetilde{\mathcal{M}_n}(Q) := \big\{ M \in \widetilde{\mathcal{M}_n} :\, \mathrm{row}_{m+1}(M) \in Q \big\}$$

and

$$\widetilde{\mathcal{M}_n}(w) := \big\{ M \in \widetilde{\mathcal{M}_n}(Q) :\, \mathrm{row}_{m+1}(M) = w \big\}, \;\; w = v, v'.$$

With these notations, we have

$$\mathbb{P}_Q(w) = \frac{|\widetilde{\mathcal{M}_n}(w)|}{|\widetilde{\mathcal{M}_n}(Q)|}, \;\; w = v, v'.$$

Next, we denote

$$\widetilde{\mathcal{M}_n}^*(w) := \big\{ M \in \widetilde{\mathcal{M}_n}(w) :\, |\mathrm{supp}\,\mathrm{col}_k(M) \cap \mathrm{supp}\,\mathrm{col}_\ell(M) \cap [m]^c| \leq 0.9d \big\}, \;\; w = v, v',$$

and for any non-negative integer $r \leq 0.9d$ we set

$$\widetilde{\mathcal{M}_n}^*(w, r) := \big\{ M \in \widetilde{\mathcal{M}_n}(w) :\, |\mathrm{supp}\,\mathrm{col}_k(M) \cap \mathrm{supp}\,\mathrm{col}_\ell(M) \cap [m]^c| = r \big\}, \;\; w = v, v'.$$

Clearly,

$$\widetilde{\mathcal{M}_n}^*(w) = \bigsqcup_{r=1}^{\lfloor 0.9d \rfloor} \widetilde{\mathcal{M}_n}^*(w, r). \tag{18}$$

Applying the "matrix" version of Proposition 2.8 to the set $\widetilde{\mathcal{M}_n}^*(v)$, we get

$$|\widetilde{\mathcal{M}_n}^*(v)| \geq \big(1 - \exp(-c_{2.8}d)\big)\, |\widetilde{\mathcal{M}_n}(v)|. \tag{19}$$

Fix an integer $r \leq 0.9d$. We shall compare the cardinalities of $\widetilde{\mathcal{M}_n}^*(v,r)$ and $\widetilde{\mathcal{M}_n}^*(v',r)$. Let us define a relation $\widetilde{R} \subset \widetilde{\mathcal{M}_n}^*(v,r) \times \widetilde{\mathcal{M}_n}^*(v',r)$ as follows:

Pick any $M \in \widetilde{\mathcal{M}_n}^*(v,r)$ and $s \in [m]^c \cap \operatorname{supp}\operatorname{col}_\ell(M) \setminus \operatorname{supp}\operatorname{col}_k(M)$. Clearly, we have $M_{(m+1)k} = M_{s\ell} = 1$ and $M_{(m+1)\ell} = M_{sk} = 0$. Let M^s be the matrix obtained from M by a simple switching operation on the entries $(m+1,k)$, $(m+1,\ell)$, (s,k), (s,ℓ). It is easy to see that M^s belongs to $\widetilde{\mathcal{M}_n}^*(v',t)$. We set $\widetilde{R}(M) := \{M^s :\ s \in [m]^c \cap \operatorname{supp}\operatorname{col}_\ell(M) \setminus \operatorname{supp}\operatorname{col}_k(M)\}$.

Thus,

$$|\widetilde{R}(M)| = |[m]^c \cap \operatorname{supp}\operatorname{col}_\ell(M) \setminus \operatorname{supp}\operatorname{col}_k(M)| = p_\ell - r.$$

Further, it is not difficult to check that $\widetilde{R}(\widetilde{\mathcal{M}_n}^*(v,r)) = \widetilde{\mathcal{M}_n}^*(v',r)$ and for any $M' \in \widetilde{\mathcal{M}_n}^*(v',r)$, we have

$$|\widetilde{R}^{-1}(M')| = |[m]^c \cap \operatorname{supp}\operatorname{col}_k(M') \setminus \operatorname{supp}\operatorname{col}_\ell(M')| = p_k - r.$$

Hence, by Claim 2.1,

$$|\widetilde{\mathcal{M}_n}^*(v,r)| = \frac{p_\ell - r}{p_k - r}\, |\widetilde{\mathcal{M}_n}^*(v',r)|.$$

Using this together with (18) and (19), we can write

$$\big(1-\exp(-c_{2.8}d)\big)\, |\widetilde{\mathcal{M}_n}(v)| \leq |\widetilde{\mathcal{M}_n}^*(v)| = \sum_{r=1}^{\lfloor 0.9d\rfloor} |\widetilde{\mathcal{M}_n}^*(v,r)| \leq \max_{1\leq r\leq 0.9d} \frac{p_\ell - r}{p_k - r}\, |\widetilde{\mathcal{M}_n}^*(v')|.$$

Finally, we divide both sides by $|\widetilde{\mathcal{M}_n}(Q)|$ and notice that

$$\max_{1\leq r\leq 0.9d} \frac{p_\ell - r}{p_k - r} = \frac{p_\ell}{p_k}\, \mathbf{1}_{p_\ell < p_k} + \Big(1 + \frac{p_\ell - p_k}{p_k - \lfloor 0.9d\rfloor}\Big)\, \mathbf{1}_{p_\ell \geq p_k}.$$

□

Remark 3.2. Note that under our assumptions on p_ℓ and d, we have

$$1 - 8c_0 \leq \gamma_{k,\ell} \leq 1 + 50c_0 \quad \text{for all } k,\ell \in \{1,2,\dots,n\}.$$

We define a relation $\mathcal{R} \subset Q \times Q$ as

$$(v,v') \in \mathcal{R} \ \text{ if and only if } \ |\{j \leq n :\ v_j \neq v'_j\}| = 2, \tag{20}$$

i.e. the pair (v,v') belongs to $\mathcal{R}$ if v' can be obtained from v by transposing two coordinates. Further, let us define sets T_+ and T_0:

$$T_+ := \{v\in Q:\ v_i = 1\} \quad \text{and} \quad T_0 := \{v\in Q:\ v_i=0\}, \tag{21}$$

so that $Q = T_+ \sqcup T_0$. Denote by $\mathcal{R}_+ \subset T_+\times T_0$ the restriction of the relation $\mathcal{R}$ to $T_+\times T_0$. For a vector $v'\in T_0$, let N be the number of coordinates of v' equal to 1, starting from the i-th coordinate. Note that this number does not depend on the choice of $v'\in T_0$, and is entirely determined by the values of the signs $\varepsilon_1,\varepsilon_2,\dots,\varepsilon_{i-1}$ which we fixed at the beginning of the sub-section. More precisely,

$$N := \mathbf{d}_i^{out} - \sum_{j<i}\varepsilon_j. \tag{22}$$

Note that, provided that both T_0, T_+ are not empty, for any $v'\in T_0$ we have $|\mathcal{R}_+^{-1}(v')| = N$. Moreover, for any $v\in T_+$, the cardinality of $\mathcal{R}_+(v)$ is the number of coordinates equal to 0 after the i-th coordinate in v. Therefore, for any $v\in T_+$, we have $|\mathcal{R}_+(v)| = n-i-N+1$.

In what follows, we will make frequent use of the quantities

$$\delta_{k,\ell} := \max\big(|1-\gamma_{k,\ell}{}^{-1}|, |1-\gamma_{\ell,k}{}^{-1}|, |1-\gamma_{k,\ell}|, |1-\gamma_{\ell,k}|\big),\quad k,\ell\in\{1,2,\dots,n\}, \tag{23}$$

where $\gamma_{k,\ell}$ are defined by (17). From Remark 3.2, it immediately follows that $\delta_{k,\ell}\le 1/4$ for all $k,\ell\in\{1,\dots,n\}$. Moreover, a simple computation shows that

$$\delta_{k,\ell} \le \frac{40}{d}|p_k - p_\ell| + 4\exp(-c_{2.8}d),\quad k,\ell\in\{1,2,\dots,n\}. \tag{24}$$

Lemma 3.3. *Suppose that the sets T_0, T_+ are non-empty. Then*

$$\begin{aligned}\big|(n-i-N+1)\mathbb{P}_Q(T_+) - N\mathbb{P}_Q(T_0)\big| &\le \mathbb{P}_Q(T_+)\frac{2(n-i-N+1)}{n-i}\sum_{\ell=i+1}^{n}\delta_{i,\ell}\\ &\le \frac12 (n-i-N+1)\mathbb{P}_Q(T_+),\end{aligned}$$

where $\delta_{k,\ell}$ are defined by (23).

Proof. First note that

$$\mathbb{P}_Q(T_+) = \sum_{v\in T_+}\mathbb{P}_Q(v) = \sum_{v\in T_+}\sum_{v'\in\mathcal{R}_+(v)}\frac{\mathbb{P}_Q(v)}{|\mathcal{R}_+(v)|} = \frac{1}{n-i-N+1}\sum_{(v,v')\in\mathcal{R}_+}\mathbb{P}_Q(v).$$

Similarly, for T_0 we have

$$\mathbb{P}_Q(T_0) = \frac1N\sum_{(v,v')\in\mathcal{R}_+}\mathbb{P}_Q(v').$$

Hence,

$$\big|(n-i-N+1)\mathbb{P}_Q(T_+) - N\mathbb{P}_Q(T_0)\big| = \Big|\sum_{(v,v')\in\mathcal{R}_+}\mathbb{P}_Q(v)-\mathbb{P}_Q(v')\Big| \le \sum_{(v,v')\in\mathcal{R}_+}\Big|1-\frac{\mathbb{P}_Q(v')}{\mathbb{P}_Q(v)}\Big|\mathbb{P}_Q(v).$$

Since for any pair $(v,v')\in\mathcal{R}_+$, v and v' differ just at one coordinate after i-th, we have

$$\mathcal{R}_+=\bigsqcup_{\ell=i+1}^{n}\big\{(v,v')\in\mathcal{R}_+:\,v_\ell\neq v'_\ell\big\},$$

whence

$$\big|(n-i-N+1)\mathbb{P}_Q(T_+)-N\mathbb{P}_Q(T_0)\big|\le\sum_{\ell=i+1}^{n}\sum_{\substack{v_\ell\neq v'_\ell\\(v,v')\in\mathcal{R}_+}}\Big|1-\frac{\mathbb{P}_Q(v')}{\mathbb{P}_Q(v)}\Big|\mathbb{P}_Q(v).$$

Applying Lemma 3.1, we obtain

$$\big|(n-i-N+1)\mathbb{P}_Q(T_+)-N\mathbb{P}_Q(T_0)\big|\le\sum_{\ell=i+1}^{n}\delta_{i,\ell}\sum_{\substack{v_\ell=0\\v\in T_+}}\mathbb{P}_Q(v).\tag{25}$$

Let us now compare the quantities $a_\ell:=\sum_{\substack{v_\ell=0\\v\in T_+}}\mathbb{P}_Q(v)$ for two different values of ℓ. Fix $\ell,\ell'>i$ ($\ell\neq\ell'$) and define a bijection $f:\{v\in T_+:\,v_\ell=0\}\to\{v\in T_+:\,v_{\ell'}=0\}$ as follows: given $v\in T_+$, if $v_\ell=v_{\ell'}=0$ then we set $f(v):=v$; otherwise, if $v_\ell=0$ and $v_{\ell'}=1$ then we let $f(v)$ be the vector obtained by swapping the ℓ-th and ℓ'-th coordinates of v. Note that whenever $v\neq f(v)$, we have $(v,f(v))\in\mathcal{R}$. Hence, using Lemma 3.1, we get

$$\frac{a_\ell}{a_{\ell'}}\le\max_{\substack{v_\ell=0\\v\in T_+}}\frac{\mathbb{P}_Q(v)}{\mathbb{P}_Q(f(v))}\le\max(1,\gamma_{\ell',\ell})\le 2,$$

where the last inequality follows from Remark 3.2. This implies

$$\max_{\ell>i}a_\ell\le 2\min_{\ell>i}a_\ell\le\frac{2}{n-i}\sum_{\ell=i+1}^{n}a_\ell.$$

Plugging this estimate into (25), we deduce that

$$\big|(n-i-N+1)\mathbb{P}_Q(T_+)-N\mathbb{P}_Q(T_0)\big|\le\frac{2}{n-i}\sum_{\ell=i+1}^{n}\delta_{i,\ell}\sum_{\ell'=i+1}^{n}a_{\ell'}.$$

The proof is finished by noticing that

$$\sum_{\ell'=i+1}^{n}a_{\ell'}=\sum_{(v,v')\in\mathcal{R}_+}\mathbb{P}_Q(v)=(n-i-N+1)\,\mathbb{P}_Q(T_+).$$

□

Assume that T_0,T_+ are non-empty. Given a couple $(v,v')\in\mathcal{R}_+$, define

$$\rho(v,v'):=\frac{\mathbb{P}_Q(v)}{(n-i-N+1)\,\mathbb{P}_Q(T_+)}\quad\text{and}\quad\rho'(v,v'):=\frac{\mathbb{P}_Q(v')}{N\,\mathbb{P}_Q(T_0)}.$$

Note that ρ and ρ' are probability measures on $\mathcal{R}_+$. In what follows, given a function $h : Q \to \mathbb{R}$, by $\mathbb{E}_{T_+}$ we denote the expectation of the restriction of h to T_+ with respect to $\mathbb{P}_Q$, i.e.,

$$\mathbb{E}_{T_+} h = \frac{1}{\mathbb{P}_Q(T_+)} \sum_{v \in T_+} h(v)\, \mathbb{P}_Q(v) = \sum_{(v,v') \in \mathcal{R}_+} \rho(v,v')\, h(v).$$

Similarly,

$$\mathbb{E}_{T_0} h = \frac{1}{\mathbb{P}_Q(T_0)} \sum_{v' \in T_0} h(v')\, \mathbb{P}_Q(v') = \sum_{(v,v') \in \mathcal{R}_+} \rho'(v,v')\, h(v').$$

We shall proceed by comparing the measures ρ and ρ':

Lemma 3.4. *Assume that the sets T_+, T_0 are non-empty. Let $(v,v') \in \mathcal{R}_+$ and let $q > i$ be an integer such that $v_q \neq v'_q$. Then*

$$|\rho(v,v') - \rho'(v,v')| \leq \rho'(v,v') \Big[\delta_{i,q} + \frac{4}{n-i} \sum_{\ell=i+1}^{n} \delta_{i,\ell} \Big].$$

Proof. Using Lemma 3.1 and the definition (23), we get

$$1 - \delta_{i,q} \leq \frac{\mathbb{P}_Q(v)}{\mathbb{P}_Q(v')} \leq 1 + \delta_{i,q}.$$

Now, from Lemma 3.3, we have

$$1 - \frac{2}{n-i} \sum_{\ell=i+1}^{n} \delta_{i,\ell} \leq \frac{N \mathbb{P}_Q(T_0)}{(n-i-N+1)\mathbb{P}_Q(T_+)} \leq 1 + \frac{2}{n-i} \sum_{\ell=i+1}^{n} \delta_{i,\ell}.$$

Recall that the assumptions on d and p_ℓ's imply that $\delta_{i,\ell} \leq 1$. Hence, putting together the last two estimates, we obtain

$$(1 - \delta_{i,q}) \Big[1 - \frac{2}{n-i} \sum_{\ell=i+1}^{n} \delta_{i,\ell} \Big] \leq \frac{\rho(v,v')}{\rho'(v,v')} \leq (1 + \delta_{i,q}) \Big[1 + \frac{2}{n-i} \sum_{\ell=i+1}^{n} \delta_{i,\ell} \Big].$$

The proof is finished by multiplying the inequalities by $\rho'(v,v')$ and employing the bound $\delta_{i,q} \leq 1$. $\square$

Lemma 3.5. *Let, as before, T_0, T_+ be given by (21), and assume that both T_0, T_+ are non-empty. Let h be any function on Q. Then for any $\lambda \in \mathbb{R}$, we have*

$$\begin{aligned} |\mathbb{E}_{T_+} h - \mathbb{E}_{T_0} h| \leq\; & \frac{1}{n-i-N+1} \sup_{v \in T_+} \sum_{v' \in \mathcal{R}_+(v)} \big| h(v) - h(v') \big| \\ & + \frac{8\, \mathbb{E}_{T_0} |h - \lambda|}{n-i} \sum_{\ell=i+1}^{n} \delta_{i,\ell} \\ & + \Big(\frac{4}{n-i} \sum_{\ell=i+1}^{n} \delta_{i,\ell} \Big) \max_{(v,v') \in \mathcal{R}} \big| h(v) - h(v') \big|, \end{aligned}$$

where N is defined by (22).

Before proving the lemma, let us comment on the idea behind the estimate. Suppose that the function h is a linear functional in $\mathbb{R}^n$ (actually this is the only case interesting for us). Then, loosely speaking, we want to show that the difference $|\mathbb{E}_{T_+}h-\mathbb{E}_{T_0}h|$ is essentially determined by the value $h(e_i)$. This corresponds to the first term of the bound, whereas the second and third summands are supposed to be negligible under appropriate conditions on h (in fact, the second summand $\frac{8\,\mathbb{E}_{T_0}|h-\lambda|}{n-i}\sum_{\ell=i+1}^n \delta_{i,\ell}$ can be problematic and requires special handling).

Proof of Lemma 3.5. Fix any $\lambda\in\mathbb{R}$. Using the triangle inequality and the definition of $\mathbb{E}_{T_+}h$ and $\mathbb{E}_{T_0}h$, we obtain

$$\begin{aligned}
\beta &:= |\mathbb{E}_{T_+}h-\mathbb{E}_{T_0}h|\\
&\leq \Big|\sum_{(v,v')\in\mathcal{R}_+}\big(h(v)-h(v')\big)\,\rho(v,v')\Big| + \Big|\sum_{(v,v')\in\mathcal{R}_+}\big(\rho(v,v')-\rho'(v,v')\big)\,h(v')\Big|\\
&= \Big|\sum_{(v,v')\in\mathcal{R}_+}\big(h(v)-h(v')\big)\,\rho(v,v')\Big| + \Big|\sum_{(v,v')\in\mathcal{R}_+}\big(\rho(v,v')-\rho'(v,v')\big)\,\big(h(v')-\lambda\big)\Big|\\
&\leq \Big|\sum_{(v,v')\in\mathcal{R}_+}\big(h(v)-h(v')\big)\,\rho(v,v')\Big| + \sum_{(v,v')\in\mathcal{R}_+}\big|\rho(v,v')-\rho'(v,v')\big|\,\big|h(v')-\lambda\big|\\
&= \Big|\sum_{(v,v')\in\mathcal{R}_+}\big(h(v)-h(v')\big)\,\rho(v,v')\Big| + \sum_{\ell=i+1}^{n}\sum_{\substack{v_\ell\neq v'_\ell\\(v,v')\in\mathcal{R}_+}}\big|\rho(v,v')-\rho'(v,v')\big|\,\big|h(v')-\lambda\big|.
\end{aligned}$$

For the first term, applying the definition of $\rho(v,v')$, we get

$$\begin{aligned}
\Big|\sum_{(v,v')\in\mathcal{R}_+}\big(h(v)-h(v')\big)\,\rho(v,v')\Big| &\leq \sum_{(v,v')\in\mathcal{R}_+}\frac{\big|h(v)-h(v')\big|\mathbb{P}_Q(v)}{(n-i-N+1)\,\mathbb{P}_Q(T_+)}\\
&\leq \frac{1}{n-i-N+1}\sup_{v\in T_+}\sum_{v'\in\mathcal{R}_+(v)}\big|h(v)-h(v')\big|.
\end{aligned}$$

Next, in view of Lemma 3.4,

$$\begin{aligned}
&\sum_{\ell=i+1}^{n}\sum_{\substack{v_\ell\neq v'_\ell\\(v,v')\in\mathcal{R}_+}}\big|\rho(v,v')-\rho'(v,v')\big|\,\big|h(v')-\lambda\big|\\
&\quad\leq \sum_{\ell=i+1}^{n}\sum_{\substack{v_\ell\neq v'_\ell\\(v,v')\in\mathcal{R}_+}}\big|h(v')-\lambda\big|\,\rho'(v,v')\Big[\delta_{i,\ell}+\frac{4}{n-i}\sum_{q=i+1}^{n}\delta_{i,q}\Big]\\
&\quad= \frac{4\mathbb{E}_{T_0}|h-\lambda|}{n-i}\sum_{\ell=i+1}^{n}\delta_{i,\ell} + \sum_{\ell=i+1}^{n}\sum_{\substack{v_\ell\neq v'_\ell\\(v,v')\in\mathcal{R}_+}}\delta_{i,\ell}\big|h(v')-\lambda\big|\,\rho'(v,v').
\end{aligned}$$

Denote

$$\alpha := \max_{(v,v')\in\mathcal{R}} |h(v)-h(v')| \quad \text{and} \quad a_\ell := \sum_{\substack{v_\ell\neq v'_\ell \\ (v,v')\in\mathcal{R}_+}} \big(\alpha+\big|h(v')-\lambda\big|\big)\,\rho'(v,v') \quad \text{for any } \ell > i.$$

Then, obviously,

$$\sum_{\ell=i+1}^{n} \sum_{\substack{v_\ell\neq v'_\ell \\ (v,v')\in\mathcal{R}_+}} \delta_{i,\ell}\big|h(v')-\lambda\big|\,\rho'(v,v') \leq \sum_{\ell=i+1}^{n} \delta_{i,\ell}\,a_\ell. \tag{26}$$

Similarly to the argument within the proof of Lemma 3.3, we shall compare a_ℓ's for any two distinct values of ℓ. Fix $\ell \neq \ell' > i$ and define a bijection $f : \{v' \in T_0 : v'_\ell = 1\} \to \{v' \in T_0 : v'_{\ell'} = 1\}$ as follows: given $v' \in T_0$ with $v'_\ell = v'_{\ell'} = 1$, set $f(v') := v'$; otherwise, if $v'_\ell = 1$ and $v'_{\ell'} = 0$ then let $f(v')$ to be the vector obtained by swapping ℓ-th and ℓ'-th coordinate of v'. Note that in the latter case $(v', f(v')) \in \mathcal{R}$. Applying Lemma 3.1, we get

$$\frac{a_\ell}{a_{\ell'}} \leq \max_{\substack{v'_\ell=1 \\ v'\in T_0}} \frac{\big(\alpha+\big|h(v')-\lambda\big|\big)\,\mathbb{P}_Q(v')}{\big(\alpha+\big|h\big(f(v')\big)-\lambda\big|\big)\,\mathbb{P}_Q(f(v'))} \leq \max(1,\gamma_{\ell,\ell'}) \max_{\substack{v'_\ell=1 \\ v'\in T_0}} \frac{\big(\alpha+\big|h(v')-\lambda\big|\big)}{\big(\alpha+\big|h\big(f(v')\big)-\lambda\big|\big)}, \tag{27}$$

where $\gamma_{\ell,\ell'}$ is defined by (17). On the other hand, since $(v', f(v')) \in \mathcal{R}$ whenever $v' \neq f(v')$, we have

$$\big|h(v')-\lambda\big| \leq \big|h\big(f(v')\big)-\lambda\big| + \big|h(v')-h\big(f(v')\big)\big| \leq \big|h\big(f(v')\big)-\lambda\big| + \alpha.$$

Plugging the last relation into (27) and using the bound $\gamma_{\ell,\ell'} \leq 2$, we get

$$a_\ell \leq 4a_{\ell'}, \quad \ell,\ell' \in \{i+1,\dots,n\}.$$

This implies

$$\max_{\ell>i} a_\ell \leq 4\min_{\ell>i} a_\ell \leq \frac{4}{n-i}\sum_{\ell>i} a_\ell = \frac{4}{n-i}\big(\alpha+\mathbb{E}_{T_0}|h-\lambda|\big).$$

Together with (26), the last relation gives

$$\sum_{\ell=i+1}^{n} \sum_{\substack{v_\ell\neq v'_\ell \\ (v,v')\in\mathcal{R}_+}} \delta_{i,\ell}\big|h(v')-\lambda\big|\,\rho'(v,v') \leq \Big(\frac{4}{n-i}\sum_{\ell=i+1}^{n}\delta_{i,\ell}\Big)\,\big(\alpha+\mathbb{E}_{T_0}|h-\lambda|\big).$$

It remains to combine the above estimates. □

Remark 3.6. We do not know if a more careful analysis can give a bound for $|\mathbb{E}_{T_+}h - \mathbb{E}_{T_0}h|$ in the above lemma, not involving dependence on $\mathbb{E}_{T_0}|h-\lambda|$.

Let, as before, h be a function on Q. We set

$$X_k := \mathbb{E}[h \mid \mathcal{F}_k \cap Q], \quad k = i-1, \dots, n,$$

where σ-algebras $\mathcal{F}_k$ are defined at the beginning of the section. Clearly, $(X_k)_{i-1 \le k \le n}$ is a martingale. Denote by $(d_k)_{i \le k \le n}$ the difference sequence, i.e.

$$d_k := X_k - X_{k-1}, \quad k = i, \dots, n.$$

Further, let M and σ_i be smallest non-negative numbers such that $|d_k| \le M$ a.s. for all $i \le k \le n$, and $\sum_{k=i+1}^n \mathbb{E}(d_k^2 \mid \mathcal{F}_{k-1} \cap Q) \le {\sigma_i}^2$ a.s. (note that, since our probability space is finite, such numbers always exist).

Lemma 3.7. *Assume that T_0 is non-empty. Then, with the above notations, we have*

$$\mathbb{E}_{T_0}(h - \mathbb{E}_{T_0} h)^2 \le {\sigma_i}^2.$$

Proof. First, note that $\mathbb{E}_{T_0}(h - \mathbb{E}_{T_0} h)^2$, viewed as a (constant) function on T_0, is just a restriction of the random variable $\mathbb{E}\big[(h - \mathbb{E}[h \mid \mathcal{F}_i \cap Q])^2 \mid \mathcal{F}_i \cap Q\big]$ to the set T_0. Hence, it is sufficient to prove the inequality

$$\mathbb{E}\big[(h - \mathbb{E}[h \mid \mathcal{F}_i \cap Q])^2 \mid \mathcal{F}_i \cap Q\big] \le {\sigma_i}^2.$$

We have

$$h - \mathbb{E}[h \mid \mathcal{F}_i \cap Q] = \sum_{k=i+1}^n d_k,$$

whence

$$\mathbb{E}\big[(h - \mathbb{E}[h \mid \mathcal{F}_i \cap Q])^2 \mid \mathcal{F}_i \cap Q\big] = \sum_{k,\ell=i+1}^n \mathbb{E}\big[d_k d_\ell \mid \mathcal{F}_i \cap Q\big] \le {\sigma_i}^2 + \sum_{\substack{k,\ell=i+1 \\ k \ne \ell}}^n \mathbb{E}\big[d_k d_\ell \mid \mathcal{F}_i \cap Q\big].$$

Finally, we note that $\mathbb{E}\big[d_k d_\ell \mid \mathcal{F}_i \cap Q\big] = 0$ for all $k \ne \ell$. $\square$

Now, we can state the main technical result of the sub-section:

Lemma 3.8. *Let, as before, the relation $\mathcal{R}$, sets T_0 and T_+ and the number N be defined by (20), (21) and (22), respectively, and let $\delta_{i,\ell}$ be given by (23). Then, with the above notation for the martingale sequence,*

$$|d_i| \le \frac{1}{n-i-N+1} \sup_{v \in T_+} \sum_{v' \in \mathcal{R}_+(v)} \big|h(v) - h(v')\big| + \frac{8\sigma_i}{n-i} \sum_{\ell=i+1}^n \delta_{i,\ell}$$
$$+ \Big(\frac{4}{n-i} \sum_{\ell=i+1}^n \delta_{i,\ell}\Big) \max_{(v,v') \in \mathcal{R}} \big|h(v) - h(v')\big|,$$

and

$$\mathbb{E}[d_i^2 \mid \mathcal{F}_{i-1}\cap Q] \le \frac{4N}{n-i-N+1}\Big[\frac{1}{n-i-N+1}\sup_{v\in T_+}\sum_{v'\in\mathcal{R}_+(v)}\big|h(v)-h(v')\big| + \frac{8\sigma_i}{n-i}\sum_{\ell=i+1}^{n}\delta_{i,\ell} + \Big(\frac{4}{n-i}\sum_{\ell=i+1}^{n}\delta_{i,\ell}\Big)\max_{(v,v')\in\mathcal{R}}\big|h(v)-h(v')\big|\Big]^2.$$

Proof. When one of the sets T_0 or T_+ is empty, we have $d_i = 0$, and the statement is obvious. Otherwise, it is easy to see that

$$X_{i-1} = \mathbb{E}_Q h = \mathbb{P}_Q(T_+)\,\mathbb{E}_{T_+}h + \mathbb{P}_Q(T_0)\,\mathbb{E}_{T_0}h \quad\text{and}\quad X_i = \mathbf{1}_{T_+}\mathbb{E}_{T_+}h + \mathbf{1}_{T_0}\mathbb{E}_{T_0}h,$$

where $\mathbf{1}_{T_+}, \mathbf{1}_{T_0}$ are indicators of the corresponding subsets of Q. Thus, we have

$$d_i = \mathbf{1}_{T_+}\mathbb{P}_Q(T_0)\,\big[\mathbb{E}_{T_+}h - \mathbb{E}_{T_0}h\big] - \mathbf{1}_{T_0}\mathbb{P}_Q(T_+)\,\big[\mathbb{E}_{T_+}h - \mathbb{E}_{T_0}h\big], \tag{28}$$

whence

$$|d_i| \le \max\big(\mathbb{P}_Q(T_+), \mathbb{P}_Q(T_0)\big)\,|\mathbb{E}_{T_+}h - \mathbb{E}_{T_0}h| \le |\mathbb{E}_{T_+}h - \mathbb{E}_{T_0}h|.$$

Applying Lemma 3.5 with $\lambda := \mathbb{E}_{T_0}h$, we get

$$|d_i| \le \frac{1}{n-i-N+1}\sup_{v\in T_+}\sum_{v'\in\mathcal{R}_+(v)}\big|h(v)-h(v')\big| + \frac{8\,\mathbb{E}_{T_0}|h-\mathbb{E}_{T_0}h|}{n-i}\sum_{\ell=i+1}^{n}\delta_{i,\ell} + \Big(\frac{4}{n-i}\sum_{\ell=i+1}^{n}\delta_{i,\ell}\Big)\max_{(v,v')\in\mathcal{R}}\big|h(v)-h(v')\big|.$$

The first part of the lemma follows by using Lemma 3.7.

Next, we calculate the conditional second moment of d_i. As an immediate consequence of (28), we get

$$d_i^2 = \mathbb{P}_Q(T_0)^2\big(\mathbb{E}_{T_+}h - \mathbb{E}_{T_0}h\big)^2\,\mathbf{1}_{T_+} + \mathbb{P}_Q(T_+)^2\big(\mathbb{E}_{T_+}h - \mathbb{E}_{T_0}h\big)^2\,\mathbf{1}_{T_0},$$

whence

$$\begin{aligned}\mathbb{E}[d_i^2 \mid \mathcal{F}_{i-1}\cap Q] &= \mathbb{P}_Q(T_0)^2\mathbb{P}_Q(T_+)\big(\mathbb{E}_{T_+}h - \mathbb{E}_{T_0}h\big)^2 + \mathbb{P}_Q(T_+)^2\mathbb{P}_Q(T_0)\big(\mathbb{E}_{T_+}h - \mathbb{E}_{T_0}h\big)^2\\ &= \mathbb{P}_Q(T_+)\mathbb{P}_Q(T_0)\,\big(\mathbb{E}_{T_+}h - \mathbb{E}_{T_0}h\big)^2.\end{aligned}$$

Applying Lemma 3.5 with $\lambda := \mathbb{E}_{T_0}h$ and Lemma 3.7, we get

$$\frac{\mathbb{E}[d_i^2 \mid \mathcal{F}_{i-1}\cap Q]}{\mathbb{P}_Q(T_+)\mathbb{P}_Q(T_0)} \le \Big[\frac{1}{n-i-N+1}\sup_{v\in T_+}\sum_{v'\in\mathcal{R}_+(v)}\big|h(v)-h(v')\big| + \frac{8\sigma_i}{n-i}\sum_{\ell=i+1}^{n}\delta_{i,\ell} + \Big(\frac{4}{n-i}\sum_{\ell=i+1}^{n}\delta_{i,\ell}\Big)\max_{(v,v')\in\mathcal{R}}\big|h(v)-h(v')\big|\Big]^2.$$

It remains to note that $\mathbb{P}_Q(T_+)\mathbb{P}_Q(T_0) \le \mathbb{P}_Q(T_+) \le \frac{2N}{n-i-N+1}$, in view of Lemma 3.3. $\square$

Both estimates of the absolute value of d_i and of its conditional variance contain the term $\frac{1}{n-i-N+1} \sup\limits_{v\in T_+} \sum\limits_{v'\in\mathcal{R}_+(v)} \big|h(v)-h(v')\big|$. In the next simple lemma, we bound the expression in the case when h is a linear functional.

Lemma 3.9. *Let a function $h : Q \to \mathbb{R}$ be given by $h(v) := \langle v, x\rangle$ for a fixed vector $x \in \mathbb{R}^n$. Further, assume that $i \le n/4$. Then*

$$\frac{1}{n-i-N+1} \sup_{v\in T_+} \sum_{v'\in\mathcal{R}_+(v)} \big|h(v)-h(v')\big| \le |x_i| + \frac{8\|x\|_1}{n}.$$

Proof. Obviously, for any couple $(v, v') \in \mathcal{R}_+$ with $v_\ell \ne v'_\ell$ for some $\ell > i$ we have

$$\big|h(v)-h(v')\big| \le |x_i| + |x_\ell|.$$

Whence, for any $v \in T_+$,

$$\sum_{v'\in\mathcal{R}_+(v)} \big|h(v)-h(v')\big| \le |\mathcal{R}_+(v)|\,|x_i| + \sum_{\ell>i:v_\ell=0} |x_\ell| \le |\mathcal{R}_+(v)|\,|x_i| + \|x\|_1.$$

It follows that

$$\frac{1}{n-i-N+1} \sup_{v\in T_+} \sum_{v'\in\mathcal{R}_+(v)} \big|h(v)-h(v')\big| < |x_i| + \frac{8\|x\|_1}{n}.$$

$\square$

3.2. $(m+1)$-st row is conditionally concentrated

In this sub-section we show that given a fixed vector $x \in \mathbb{R}^n$ and a random vector v distributed on Ω according to the measure $\mathbb{P}_\Omega$, the scalar product $\langle v, x\rangle$ is concentrated around its expectation. Naturally, this holds under some extra assumptions on the quantities p_ℓ introduced at the beginning of the section, which measure how close to "homogeneous" the probability space $(\Omega, \mathbb{P}_\Omega)$ is. As everywhere in the sub-section, we assume that the degree sequences and parameters p_ℓ satisfy conditions (6) and (16). Additionally, throughout the sub-section we assume that

$$d \ge C_{3.2} \ln^2 n, \tag{29}$$

where $C_{3.2}$ is a sufficiently large universal constant (let us note that in its full strength the assumption is only used in the proof of Lemma 3.11 below). Define a vector $\mathcal{P} = (\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_n)$ as

$$\mathcal{P}_\ell := \sum_{j=1}^n |p_\ell - p_j|, \quad \ell \le n.$$

Note that, in view of (24) and (29), we have

$$\sum_{\ell=1}^{n} \delta_{i,\ell} \leq \frac{40}{d}\mathcal{P}_i + 1 \tag{30}$$

for any $i \leq n$.

In the previous sub-section, we estimated parameters of the martingale difference sequence generated by the variable $\langle \cdot, x\rangle$ and σ-algebras $\mathcal{F}_\ell$. Recall that the estimate of the upper bound for $|d_i|$ from Lemma 3.8 involves the quantity $\frac{\sigma_i}{n-i}\sum_{\ell=1}^n \delta_{i,\ell}$. In Section 4, applying (30), we will show that for "most" indices i, the sum $\sum_{\ell=1}^n \delta_{i,\ell}$ is bounded by $O(n/\sqrt{d})$, whereas, as we shall see below, $\sigma_i = O(\sqrt{d/n})$ for any unit vector x. Thus, the magnitude of $\frac{\sigma_i}{n-i}\sum_{\ell=1}^n \delta_{i,\ell}$ is of order $n^{-1/2}$, and it is necessarily dominated by a constant multiple of $\|x\|_\infty$. However, for *some* indices i the sum $\sum_{\ell=1}^n \delta_{i,\ell}$ can be as large as $n\ln n/\sqrt{d}$. Thus, a straighforward argument would give $C(\|x\|_\infty + n^{-1/2}\ln n)$ as an upper bound for d_i, and the implied row concentration inequality would bear the logarithmic error term. To overcome this problem, we have to consider separately two cases: when the $\|\cdot\|_\infty$-norm of the vector x is "large" and when it is "small". In the first case (treated in Lemma 3.10) the logarithmic spikes of the vector $\mathcal{P}$ do not create problems. In the second case, however, we have to apply a special ordering to coordinates of the row so that large spikes of $\mathcal{P}$ are "balanced" by a small magnitude of σ_i (which, for those coordinates i, must be much smaller than $\sqrt{d/n}$). The second case is more technically involved and is given in Lemma 3.11. Finally, when we have both statements in possession, we can complete the proof of the row concentration inequality.

Lemma 3.10. *For any $L > 0$ there exist $\alpha = \alpha(L) \in (0, 1)$ and $\beta = \beta(L) \in (0, 1)$ with the following property. Let $x \in S^{n-1}$ be an αn-sparse vector and assume that 1) $\|x\|_\infty \geq \ln(2n)\, n^{-1/2}$ and 2) $\|\mathcal{P}\|_{\psi,n} \leq Ln\sqrt{d}$, where the norm $\|\cdot\|_{\psi,n}$ is defined by (1). Then, denoting by η the random variable*

$$\eta = \eta(v) := \langle v, x\rangle - \mathbb{E}_\Omega\langle \cdot, x\rangle, \quad v \in \Omega,$$

we have

$$\mathbb{E}_\Omega\, e^{\beta\lambda\eta} \leq \exp\left(\frac{d}{n\|x\|_\infty^2}\, g(\lambda\|x\|_\infty)\right), \quad \lambda > 0,$$

with $g(\cdot)$ defined by (9).

Proof. Let $L > 0$ be fixed. We define $\alpha = \alpha(L)$ as the largest number in $(0, 1/4]$ such that

$$32 \cdot 640 C_{2.2} L\sqrt{\alpha}\ln\frac{2}{\alpha} \leq 1, \tag{31}$$

where the constant $C_{2.2}$ is given in Lemma 2.2.

Pick an αn-sparse vector $x \in S^{n-1}$ and let π be a permutation on $[n]$ such that $|x_{\pi(1)}| \geq |x_{\pi(2)}| \geq \ldots \geq |x_{\pi(n)}|$. For any $i \leq n$, we denote by $\pi(\mathcal{F}_i)$ the

σ-algebra generated by coordinates $\pi(1),\dots,\pi(i)$ of a vector distributed on Ω according to the measure $\mathbb{P}_\Omega$, i.e.

$$\pi(\mathcal{F}_i):=\sigma\big(\big\{v\in\Omega:\ v_{\pi(j)}=b_j \text{ for all } j\le i\big\},\ b_1,b_2,\dots,b_i\in\{0,1\}\big).$$

Define a function h on Ω by

$$h(v):=\langle v,x\rangle,\quad v\in\Omega,$$

and let

$$X_\ell:=\mathbb{E}_\Omega[h\mid\pi(\mathcal{F}_\ell)],\ \ \ell\le n,$$

and $d_\ell:=X_\ell-X_{\ell-1}$. Further, let M and σ be the smallest non-negative numbers such that $|d_\ell|\le M$ everywhere on Ω for all $\ell\le n$, and $\sum_{\ell=1}^n\mathbb{E}(d_\ell^2\,|\,\pi(\mathcal{F}_{\ell-1}))\le\sigma^2$ everywhere on Ω. Clearly, for any $i>\alpha n$ we have $d_i=0$. Now, fix $i\le\alpha n$ and follow the notations of the previous sub-section (with $\pi(\ell)$ replacing ℓ where appropriate). More precisely, we take an atom of the algebra $\pi(\mathcal{F}_{i-1})$ i.e. the set Q of vectors in Ω with some prescribed values of their coordinates with indices $\pi(1),\dots,\pi(i-1)$. Then $\mathcal{R}$ is a collection of all pairs of vectors from Q which differ by two coordinates and N is the number of non-zero coordinates in every $v\in Q$, excluding coordinates with indices $\pi(1),\dots,\pi(i-1)$. In view of the choice of π and the definition of h, we have

$$\max_{(v,v')\in\mathcal{R}}|h(v)-h(v')|\le 2|x_{\pi(i)}|.$$

Further, using the condition $\delta_{\pi(i),\ell}\le 1/4$, we get

$$\Big(\frac{4}{n-i}\sum_{\ell=1}^n\delta_{\pi(i),\ell}\Big)\max_{(v,v')\in\mathcal{R}}\big|h(v)-h(v')\big|\le 4|x_{\pi(i)}|.$$

Together with Lemma 3.8, Lemma 3.9 and (30), this gives

$$|d_i|\le 5|x_{\pi(i)}|+\frac{8\|x\|_1}{n}+\frac{640\sigma}{dn}\mathcal{P}_{\pi(i)}+\frac{16\sigma}{n}$$

everywhere on Q and, in fact, everywhere on Ω as the right-hand side of the last relation does not depend on the choice of atom Q. Further, applying the second part of Lemma 3.8 with Lemma 3.9 and relations $i\le n/4$, $N\le d\le n/2+c_0n$ and (30), we get

$$\begin{aligned}\mathbb{E}[d_i^2\mid\mathcal{F}_{i-1}]&\le\frac{32d}{n}\Big[5|x_{\pi(i)}|+\frac{8\|x\|_1}{n}+\frac{16\sigma}{n}\sum_{\ell=1}^n\delta_{\pi(i),\ell}\Big]^2\\&\le\frac{32d}{n}\Big[75{x_{\pi(i)}}^2+\frac{48}{n}+3\Big(\frac{640\sigma}{dn}\mathcal{P}_{\pi(i)}+\frac{16\sigma}{n}\Big)^2\Big],\end{aligned}$$

where in the last inequality we used the convexity of the square and $\|x\|_1^2\le\alpha n\,\|x\|_2^2\le n/4$. Again, the bound for $\mathbb{E}[d_i^2\mid\mathcal{F}_{i-1}]$ holds everywhere on Ω.

Summing over all $i \leq \alpha n$, we get from the last relation

$$\sigma^2 \leq \frac{32d}{n}\Big[87 + 3\sum_{i=1}^{\lfloor \alpha n \rfloor}\Big(\frac{640\sigma}{dn}\mathcal{P}_{\pi(i)} + \frac{16\sigma}{n}\Big)^2\Big]$$
$$\leq \frac{32d}{n}\Big[87 + \frac{6\cdot 640^2\sigma^2}{d^2n^2}\sum_{i=1}^{\lfloor \alpha n \rfloor}{\mathcal{P}_{\pi(i)}}^2 + \frac{6\cdot 16^2\sigma^2}{n}\Big].$$

In view of the condition on $\|\mathcal{P}\|_{\psi,n}$, relation (31) and Lemma 2.2, we have

$$\frac{32d}{n}\frac{6\cdot 640^2\sigma^2}{d^2n^2}\sum_{i=1}^{\lfloor \alpha n \rfloor}{\mathcal{P}_{\pi(i)}}^2 \leq \frac{\sigma^2}{4}.$$

Thus, the self-bounding estimate for σ implies

$$\sigma^2 < \frac{Cd}{n},$$

for an appropriate constant $C > 0$. Whence, from the above estimate of $|d_i|$'s we obtain

$$M \leq 5\|x\|_\infty + \frac{8\|x\|_1}{n} + \frac{640\sigma}{dn}\|\mathcal{P}\|_\infty + \frac{16\sigma}{n}$$
$$\leq 5\|x\|_\infty + \frac{8}{\sqrt{n}} + \frac{640\sqrt{C}\,L\ln(en)}{\sqrt{n}} + \frac{16\sqrt{C}}{n}$$

where we employed the relations $\|\mathcal{P}\|_\infty \leq \ln(en)\|\mathcal{P}\|_{\psi,n} \leq Ln\sqrt{d}\ln(en)$ (see formula (7)) and the estimate for σ established above. This, together with the assumption $\|x\|_\infty \geq \ln(2n)\,n^{-1/2}$, implies that $M \leq C'(1+L)\,\|x\|_\infty$ for an appropriate constant C'. It remains to apply Theorem 2.5 in order to finish the proof. □

The next lemma is a counterpart of the above statement, covering the case when the $\|\cdot\|_\infty$-norm of the vector x is small.

Lemma 3.11. *For any $L > 0$ there exist $\alpha = \alpha(L) \in (0,1)$ and $\beta = \beta(L) \in (0,1)$ with the following property. Let $x \in S^{n-1}$ be an αn-sparse vector and assume that 1) $\|x\|_\infty < \ln(2n)\,n^{-1/2}$ and 2) $\|\mathcal{P}\|_{\psi,n} \leq Ln\sqrt{d}$. Then, denoting by η the random variable*

$$\eta = \eta(v) := \langle v, x\rangle - \mathbb{E}_\Omega\langle\cdot, x\rangle, \quad v \in \Omega,$$

we have

$$\mathbb{E}_\Omega\, e^{\beta\lambda\eta} \leq \exp\Big(\frac{d}{n\|x\|_\infty^2}\,g(\lambda\|x\|_\infty)\Big), \quad \lambda > 0.$$

Proof. Again, we fix $L > 0$. Let $C_1 > 0$ be a large enough universal constant (whose exact value can be determined from the proof below). We define $\alpha = \alpha(L)$ as the largest number in $(0, 1/4]$ such that

$$C_1 L^2 \alpha \ln^2 \frac{e}{\alpha} \le \frac{1}{2}. \tag{32}$$

Let π be a permutation on $[n]$ such that $|x_{\pi(i)}| = 0$ for all $i > |\mathrm{supp}\, x|$ and the sequence $(\mathcal{P}_{\pi(i)})_{i \le |\mathrm{supp}\, x|}$ is non-decreasing. We define the function h, σ-algebras $\pi(\mathcal{F}_\ell)$ and the difference sequence $(d_\ell)_{\ell \le n}$ the same way as in the proof of Lemma 3.10. We have $d_i = 0$ for all $i > |\mathrm{supp}\, x|$. Let M and σ_i ($i \le |\mathrm{supp}\, x|$) be the smallest numbers such that everywhere on Ω we have $|d_i| \le M$ for all $i \le |\mathrm{supp}\, x|$ and

$$\sum_{\ell=i}^{|\mathrm{supp}\, x|} \mathbb{E}(d_\ell^2 \,|\, \pi(\mathcal{F}_{\ell-1})) \le {\sigma_i}^2, \quad i \le |\mathrm{supp}\, x|.$$

We fix any $i \le |\mathrm{supp}\, x|$ and follow the notations of the previous sub-section (the way it was described in Lemma 3.10). Recall that $\max_{(v,v') \in \mathcal{R}} \big|h(v) - h(v')\big| \le 2\|x\|_\infty$. Now, using Lemmas 3.8 and 3.9, inequality (30), as well as relations $i \le n/4$ and $N \le d \le n/2 + c_0 n$, we obtain

$$\begin{aligned}
\mathbb{E}[d_i^2 \mid \pi(\mathcal{F}_{i-1})] &\le \frac{32d}{n}\Big[|x_{\pi(i)}| + \frac{8\|x\|_1}{n} + (\|x\|_\infty + \sigma_i)\frac{8}{n-i}\sum_{\ell=1}^{n} \delta_{\pi(i),\ell}\Big]^2 \\
&\le \frac{32d}{n}\Big[|x_{\pi(i)}| + \frac{8}{\sqrt{n}} + (\|x\|_\infty + \sigma_i)\frac{16}{n}\Big(\frac{40\mathcal{P}_{\pi(i)}}{d} + 1\Big)\Big]^2 \\
&\le \frac{32d}{n}\Big[4{x_{\pi(i)}}^2 + \frac{256}{n} + (4\|x\|_\infty^2 + 4\sigma_i^2)\frac{256}{n^2}\Big(\frac{40\mathcal{P}_{\pi(i)}}{d} + 1\Big)^2\Big],
\end{aligned}$$

where in the last inequality we used the convexity of the square. Since $\sigma_\ell \le \sigma_i$ for any $i \le \ell \le |\mathrm{supp}\, x|$, we have

$$\mathbb{E}[d_\ell^2 \mid \pi(\mathcal{F}_{\ell-1})] \le \frac{32d}{n}\Big[4{x_{\pi(\ell)}}^2 + \frac{256}{n} + (4\|x\|_\infty^2 + 4\sigma_i^2)\frac{256}{n^2}\Big(\frac{40\mathcal{P}_{\pi(\ell)}}{d} + 1\Big)^2\Big]$$

for any $i \le \ell \le |\mathrm{supp}\, x|$. Summing over all such ℓ's, we get

$${\sigma_i}^2 \le \frac{128d}{n}\Big[\sum_{\ell=i}^{|\mathrm{supp}\, x|} {x_{\pi(\ell)}}^2 + \frac{64}{n}\big(|\mathrm{supp}\, x| - i + 1\big) + (\|x\|_\infty^2 + \sigma_i^2)\frac{256}{n^2}\sum_{\ell=i}^{|\mathrm{supp}\, x|}\Big(\frac{40\mathcal{P}_{\pi(\ell)}}{d} + 1\Big)^2\Big]. \tag{33}$$

Note that, by the definition of $\|\cdot\|_{\psi,n}$-norm and in view of the fact that the sequence $(\mathcal{P}_{\pi(\ell)})_{\ell \le |\mathrm{supp}\, x|}$ is non-decreasing, we get

$$\mathcal{P}_{\pi(\ell)} \le Ln\sqrt{d} \ln\left(\frac{en}{|\mathrm{supp}\, x| - \ell + 1}\right), \quad i \le \ell \le |\mathrm{supp}\, x|. \tag{34}$$

Moreover, Lemma 2.2 implies

$$\sum_{\ell=i}^{|\mathrm{supp}\, x|} {\mathcal{P}_{\pi(\ell)}}^2 \le CL^2 n^2 d\big(|\mathrm{supp}\, x| - i + 1\big) \ln^2\left(\frac{en}{|\mathrm{supp}\, x| - i + 1}\right),$$

for a sufficiently large universal constant C. Plugging in the estimate into (33), we get

$$\sigma_i{}^2 \le \frac{C'd}{n} \sum_{\ell=i}^{|\mathrm{supp}\, x|} {x_{\pi(\ell)}}^2 + \frac{C'dm}{n^2} + C'L^2(\|x\|_\infty^2 + \sigma_i{}^2)\frac{m}{n}\ln^2\left(\frac{en}{m}\right),$$

for an appropriate constant C', where $m := |\mathrm{supp}\, x| - i + 1$. Now, if C_1 in (32) is sufficiently large, the above self-bounding estimate for σ_i implies

$$\sigma_i{}^2 \le \frac{2C_1 d}{n} \sum_{\ell=i}^{|\mathrm{supp}\, x|} {x_{\pi(\ell)}}^2 + \frac{2C_1 dm}{n^2} + 2C_1 L^2\|x\|_\infty^2 \frac{m}{n}\ln^2\left(\frac{en}{m}\right).$$

Using the condition $\|x\|_\infty \le \ln(2n)/\sqrt{n}$, the assumption on d given by (29) and relation (32), we obtain

$$\sigma^2 := \sigma_1{}^2 \le C_2\frac{d}{n},$$

for an appropriate constant C_2 and

$$\sigma_i{}^2 \le (1+L^2)C_3 d\|x\|_\infty^2 \frac{|\mathrm{supp}\, x| - i + 1}{n}, \quad i \le |\mathrm{supp}\, x|,$$

for a sufficiently large constant C_3.

Now, let us turn to estimating the absolute value of d_i's. Again, we fix any $i \le |\mathrm{supp}\, x|$ and follow notations of the previous sub-section, replacing ℓ with $\pi(\ell)$ where appropriate. By Lemmas 3.8 and 3.9, inequality (30) and the above estimate of σ_i, we have

$$\begin{aligned} |d_i| &\le |x_{\pi(i)}| + \frac{8\|x\|_1}{n} + (\|x\|_\infty + \sigma_i)\frac{8}{n-i}\sum_{\ell=1}^{n}\delta_{\pi(i),\ell} \\ &\le C_4\|x\|_\infty\Big[1 + \frac{L}{n\sqrt{d}}\sqrt{\frac{|\mathrm{supp}\, x| - i + 1}{n}}\mathcal{P}_{\pi(i)}\Big], \end{aligned}$$

for some constant $C_4 > 0$. Using first (34) then the relation (32), we deduce that

$$|d_i| \le C_4(1+L)\|x\|_\infty.$$

Thus, we get that $M \le C_4(1+L)\|x\|_\infty$. Finally, we apply Theorem 2.5 with parameters M and σ estimated above. □

Now, we can state the main result of the section.

Theorem 3.12. *For any $L > 0$ there is $\gamma(L) \in (0, 1]$ with the following property: Assume that $\|\mathcal{P}\|_{\psi,n} \le Ln\sqrt{d}$, let $x \in S^{n-1}$, and denote by η the random variable*

$$\eta = \eta(v) := \langle v, x\rangle - \mathbb{E}_\Omega\langle x, \cdot\rangle, \quad v \in \Omega.$$

Then

$$\mathbb{E}_\Omega\, e^{\gamma\lambda\eta} \le \exp\left(\frac{d}{n\|x\|_\infty^2}\, g(\lambda\|x\|_\infty)\right), \quad \lambda > 0.$$

Proof. Let $\alpha = \alpha(L) \in (0, 1)$ be the largest number in $(0, 1/4]$ satisfying both (31) and (32). We represent the vector x as a sum

$$x = x^1 + x^2 + \ldots + x^m,$$

where $x^1, x^2, \ldots, x^m$ are vectors with pairwise disjoint supports such that $|\mathrm{supp}\, x^j| \le \alpha n$ $(j \le m)$ and $m := \lceil n/\lfloor \alpha n\rfloor\rceil$. For every $j \le m$, applying either Lemma 3.10 or Lemma 3.11 (depending on the $\|\cdot\|_\infty$-norm of $x^j/\|x^j\|_2$), we obtain

$$\max\left(\mathbb{E}e^{\beta\lambda\eta_j}, \mathbb{E}e^{-\beta\lambda\eta_j}\right) \le \exp\left(\frac{d\|x^j\|_2^2}{n\|x^j\|_\infty^2}\, g(\lambda\|x^j\|_\infty)\right) \le \exp\left(\frac{d}{n\|x\|_\infty^2}\, g(\lambda\|x\|_\infty)\right), \quad \lambda > 0,$$

for some $\beta = \beta(L) > 0$, where

$$\eta_j := \langle x^j, v\rangle - \mathbb{E}_\Omega\langle x^j, \cdot\rangle, \quad v \in \Omega.$$

Since $\eta = \eta_1 + \eta_2 + \ldots + \eta_m$ everywhere on Ω, we get from Hölder's inequality

$$\mathbb{E}e^{\beta\lambda\eta} = \mathbb{E}\prod_{j=1}^m e^{\beta\lambda\eta_j} \le \left(\prod_{j=1}^m \mathbb{E}e^{\beta m\lambda\eta_j}\right)^{\frac{1}{m}} \le \exp\left(\frac{d}{n\|x\|_\infty^2}\, g(\lambda m\|x\|_\infty)\right).$$

The statement follows with $\gamma := \beta/m$. □

The above theorem leaves open the question of estimating the expectation $\mathbb{E}_\Omega\langle\cdot, x\rangle$. This problem is addressed in the last statement of the section.

Proposition 3.13. *For any non-zero vector $x \in \mathbb{R}^n$ we have*

$$\left|\mathbb{E}_\Omega\langle\cdot, x\rangle - \frac{\mathbf{d}_{m+1}^{out}}{n}\sum_{i=1}^n x_i\right| \le \frac{C_{3.13} d\|x\|_1}{n^2} + \frac{C_{3.13}}{n}\|x\|_{\log,n}\|\mathcal{P}\|_{\psi,n}$$

where $C_{3.13} > 0$ is a sufficiently large universal constant and $\|\cdot\|_{\log,n}$ is defined by (2).

Proof. Let $\mathbf{V}$ be a random vector distributed on Ω according to the measure $\mathbb{P}_\Omega$. First, we compare expectations of individual coordinates of $\mathbf{V}$, using Lemma 3.1. We let $\gamma_{i,j}$ be defined by (17). Recall that according to Remark 3.2, we have $1 - 8c_0 \le \gamma_{i,j} \le 1 + 50c_0$. Take any $i \ne j \le n$ and define a bijective map $f : \Omega \to \Omega$ as

$$f\big((v_1, v_2, \ldots, v_n)\big) := (v_{\sigma(1)}, v_{\sigma(2)}, \ldots, v_{\sigma(n)}), \ \ (v_1, v_2, \ldots, v_n) \in \Omega,$$

where σ is the transposition of i and j. Then for any $v \in \Omega$, in view of Lemma 3.1, we have

$$\mathbb{P}_\Omega(v) \le \max\bigl(\gamma_{i,j}, \gamma_{j,i}\bigr)\mathbb{P}_\Omega(f(v)).$$

Hence,

$$\begin{aligned}\mathbb{E}_\Omega \mathbf{V}_i &= \mathbb{P}_\Omega\{v \in \Omega :\ v_i = 1\}\\ &\le \max\bigl(\gamma_{i,j}, \gamma_{j,i}\bigr) \sum_{v\in\Omega: v_i=1} \mathbb{P}_\Omega\bigl(f(v)\bigr)\\ &= \max\bigl(\gamma_{i,j}, \gamma_{j,i}\bigr)\mathbb{P}_\Omega\{v \in \Omega :\ v_j = 1\}\\ &= \max\bigl(\gamma_{i,j}, \gamma_{j,i}\bigr)\mathbb{E}_\Omega \mathbf{V}_j.\end{aligned}$$

Together with an obvious relation $\sum_{i=1}^n \mathbb{E}_\Omega \mathbf{V}_i = \mathbf{d}^{out}_{m+1}$, this implies for any fixed $i \le n$:

$$\sum_{j=1}^n \max\bigl(\gamma_{i,j}, \gamma_{j,i}\bigr)^{-1}\mathbb{E}_\Omega \mathbf{V}_i \le \mathbf{d}^{out}_{m+1} \le \sum_{j=1}^n \max\bigl(\gamma_{i,j}, \gamma_{j,i}\bigr)\mathbb{E}_\Omega \mathbf{V}_i,$$

whence

$$\Bigl|\mathbb{E}_\Omega \mathbf{V}_i - \frac{\mathbf{d}^{out}_{m+1}}{n}\Bigr| \le \frac{C\mathbf{d}^{out}_{m+1}}{n}\Bigl(\frac{1}{n}\sum_{j=1}^n \delta_{i,j}\Bigr),$$

where $\delta_{i,j}$ are defined by (23) and $C > 0$ is a universal constant.

Thus, for any non-zero vector $x = (x_1, x_2, \dots, x_n)$ we get, in view of (30),

$$\begin{aligned}\Bigl|\mathbb{E}_\Omega\langle \mathbf{V}, x\rangle - \frac{\mathbf{d}^{out}_{m+1}}{n}\sum_{i=1}^n x_i\Bigr| &\le \frac{C\mathbf{d}^{out}_{m+1}}{n}\sum_{i=1}^n |x_i|\Bigl(\frac{1}{n}\sum_{j=1}^n \delta_{i,j}\Bigr)\\ &\le \frac{C'\mathbf{d}^{out}_{m+1}}{n}\sum_{i=1}^n |x_i|\Bigl(\frac{1}{nd}\mathcal{P}_i + \frac{1}{n}\Bigr)\\ &= \frac{C'\mathbf{d}^{out}_{m+1}\|x\|_1}{n^2} + \frac{C'\mathbf{d}^{out}_{m+1}}{n^2 d}\sum_{i=1}^n |x_i|\mathcal{P}_i,\end{aligned}$$

where C' is a universal constant. Finally, applying Fenchel's inequality to the sum on the right hand side and using the definition of the Orlicz norms $\|\cdot\|_{\psi,n}$ and $\|\cdot\|_{\log,n}$, we obtain

$$\begin{aligned}\sum_{i=1}^n |x_i|\mathcal{P}_i &= \|x\|_{\log,n}\|\mathcal{P}\|_{\psi,n}\sum_{i=1}^n \frac{|x_i|}{\|x\|_{\log,n}}\,\frac{\mathcal{P}_i}{\|\mathcal{P}\|_{\psi,n}}\\ &\le \|x\|_{\log,n}\|\mathcal{P}\|_{\psi,n}\sum_{i=1}^n \left(\frac{|x_i|}{\|x\|_{\log,n}}\ln_+\left(\frac{|x_i|}{\|x\|_{\log,n}}\right) + \exp\left(\mathcal{P}_i/\|\mathcal{P}\|_{\psi,n}\right)\right)\\ &\le (e+1)n\,\|x\|_{\log,n}\|\mathcal{P}\|_{\psi,n}.\end{aligned}$$

The result follows. □

4. Tensorization

The goal of this section is to transfer the concentration inequality for a single row obtained in the previous section (Theorem 3.12) to the whole matrix. Throughout the section, we assume that the degree sequences $\mathbf{d}^{in}$, $\mathbf{d}^{out}$ satisfy (6) for some d, and that d itself satisfies (29). Moreover, we always assume that the set of matrices $\mathcal{M}_n(\mathbf{d}^{in}, \mathbf{d}^{out})$ is non-empty. It will be convenient to introduce in this section a "global" random object — a matrix $\mathbf{M}$ uniformly distributed on $\mathcal{M}_n(\mathbf{d}^{in}, \mathbf{d}^{out})$.

4.1. Edge count statistics

In this subsection, the aim will be to show that the event given by (4) holds with high probability. Let G be a directed graph on n vertices with degree sequences $\mathbf{d}^{in}$, $\mathbf{d}^{out}$, and let $M = (M_{ij})$ be the adjacency matrix of G. Next, let I be a subset of $[n]$ (possibly, empty). We define quantities $p_j^{col}(I, M)$, $p_j^{row}(I, M)$ $(j \leq n)$ as in the Introduction (let us repeat the definition here for convenience):

$$p_j^{col}(I, M) := \mathbf{d}_j^{in} - |\{q \in I :\ M_{qj} = 1\}| = |\{q \in I^c :\ M_{qj} = 1\}|;$$
$$p_j^{row}(I, M) := \mathbf{d}_j^{out} - |\{q \in I :\ M_{jq} = 1\}| = |\{q \in I^c :\ M_{jq} = 1\}|.$$

Again, we define vectors $\mathcal{P}^{col}(I, M)$, $\mathcal{P}^{row}(I, M) \in \mathbb{R}^n$ coordinate-wise as

$$\mathcal{P}_j^{col}(I, M) := \sum_{\ell=1}^{n} |p_j^{col}(I, M) - p_\ell^{col}(I, M)|;$$
$$\mathcal{P}_j^{row}(I, M) := \sum_{\ell=1}^{n} |p_j^{row}(I, M) - p_\ell^{row}(I, M)|.$$

Clearly, these objects are close relatives of the quantities p_j and the vector $\mathcal{P}$ defined in the previous section. In fact, if $\widetilde{\mathcal{M}_n}$ is the subset of all matrices from $\mathcal{M}_n(\mathbf{d}^{in}, \mathbf{d}^{out})$ with a fixed realization of rows from I then $p_j^{col}(I, \cdot)$ $(j \leq n)$ and $\mathcal{P}^{col}(I, \cdot)$ are constants on $\widetilde{\mathcal{M}_n}$, which, up to relabelling the graph vertices, correspond to p_j's and $\mathcal{P}$ from Section 3.

Note that Theorem 3.12 operates under assumption that the vector $\mathcal{P}$, or, in context of this section, random vectors $\mathcal{P}^{col}(I, \mathbf{M})$ for appropriate subsets I, have small magnitude in $\|\cdot\|_{\psi,n}$-norm — the fact which still needs to be established. For any $L > 0$, let $\mathcal{E}_{\mathcal{P}}(L)$ be given by (4), i.e.

$$\mathcal{E}_{\mathcal{P}}(L) = \Big\{ \|\mathcal{P}^{row}(I, \mathbf{M})\|_{\psi,n}, \|\mathcal{P}^{col}(I, \mathbf{M})\|_{\psi,n} \leq Ln\sqrt{d}$$
$$\text{for any interval subset } I \subset [n] \text{ of cardinality at most } c_0 n \Big\}.$$

To make Theorem 3.12 useful, we need to show that for some appropriately chosen parameter L the event $\mathcal{E}_{\mathcal{P}}(L)$ has probability close to one. Obviously, this will require much stronger assumptions on the degree sequences than ones we employed up to this point. But, even under the stronger assumptions on $\mathbf{d}^{in}, \mathbf{d}^{out}$, proving an upper estimate for $\|\mathcal{P}^{col}(I,\mathbf{M})\|_{\psi,n}$, $\|\mathcal{P}^{row}(I,\mathbf{M})\|_{\psi,n}$ will require us to use the concentration results from Section 3. In order not to create a vicious cycle, we will argue in the following manner: First, we apply Theorem 3.12 in the situation when the set I has very small cardinality. It can be shown that in this case we get the required assumptions on $\|\mathcal{P}^{col}(I,\mathbf{M})\|_{\psi,n}$ for free, as long as the degree sequences satisfy certain additional conditions. This, in turn, will allow us to establish the required bounds for $\|\mathcal{P}^{col}(I,\mathbf{M})\|_{\psi,n}$ for "large" subsets I. Finally, having this result in possession, we will be able to use the full strength of Theorem 3.12 and complete the tensorization.

Let us note that condition $\|\mathcal{P}^{col}(I,M)\|_{\psi,n} = O(n\sqrt{d})$ for a matrix $M \in \mathcal{M}_n(\mathbf{d}^{in}, \mathbf{d}^{out})$ and a subset I of cardinality at most $c_0 n$ automatically implies an analog of condition (16), as long as n is sufficiently large. To be more precise, we have the following

Lemma 4.1. *There is a universal constant $c_{4.1} > 0$ with the following property: Assume that for some matrix $M \in \mathcal{M}_n(\mathbf{d}^{in}, \mathbf{d}^{out})$ and $I \subset [n]$ with $|I| \le c_0 n$ we have*

$$\|\mathcal{P}^{col}(I,M)\|_{\psi,n} \le c_{4.1} nd/\ln n.$$

Then necessarily

$$\mathbf{d}^{in}_j \ge p^{col}_j(I,M) \ge (1-2c_0)\mathbf{d}^{in}_j$$

for all $j \le n$.

Proof. Assume that $p^{col}_i(I,M) < (1-2c_0)\mathbf{d}^{in}_i$ for some $i \le n$. Define

$$J := \big\{ j \le n : p^{col}_j(I,M) < (1-1.5c_0)\mathbf{d}^{in}_j \big\}.$$

Then, obviously,

$$\big|\big\{(k,\ell) \in I \times [n] :\, M_{k\ell} = 1\big\}\big| \ge 1.5c_0 \sum_{j\in J} \mathbf{d}^{in}_j \ge 1.4 c_0 d|J|.$$

On the other hand,

$$\big|\big\{(k,\ell) \in I \times [n] :\, M_{k\ell} = 1\big\}\big| = \sum_{k\in I} \mathbf{d}^{out}_k \le c_0 nd.$$

Thus, $|J| \le \frac{5}{7}n$. This implies that

$$\mathcal{P}^{col}_i(I,M) \ge \sum_{k\in J^c} \big|p^{col}_i(I,M) - p^{col}_k(I,M)\big| > c_0|J^c|d/4 \ge c_0 nd/14.$$

Hence, by (7), we get

$$\|\mathcal{P}^{col}(I,M)\|_{\psi,n} > \frac{c_0 nd}{14\ln(en)}.$$

The result follows. □

The above lemma allows us not to worry about condition (16) and focus our attention on the $\|\cdot\|_{\psi,n}$-norm of vectors $\mathcal{P}^{col}(I,\mathbf{M})$. The bounds for $\|\mathcal{P}^{col}(I,\mathbf{M})\|_{\psi,n}$ are obtained in Proposition 4.5. But first we need to consider two auxiliary statements.

Lemma 4.2. *For any $L>0$ there are $\gamma(L)\in(0,1]$ and $K=K(L)>0$ such that the following holds. Let the degree sequences $\mathbf{d}^{in}$ and $\mathbf{d}^{out}$ be such that $\big\|(\mathbf{d}_i^{in}-d)_{i=1}^n\big\|_{\psi,n}, \big\|(\mathbf{d}_i^{out}-d)_{i=1}^n\big\|_{\psi,n}\le L\sqrt{d}$, where $\|\cdot\|_{\psi,n}$ is defined by (1). Further, let $J\subset[n]$ be a subset of cardinality $\sqrt{d}/2\le|J|\le\sqrt{d}$, and let $I\subset[n]$ be any non-empty subset. Define a $|J|$-dimensional random vector in $\mathbb{R}^J$ as*

$$v(I):=(v_k)_{k\in J},\quad v_k:=\Big|p_k^{row}(I,\mathbf{M})-\frac{\mathbf{d}_k^{out}|I^c|}{n}\Big|,\ k\in J.$$

Then for any subset $T\subset J$ and any $t\ge K\sqrt{d}\,|T|$, we have

$$\mathbb{P}\Big\{\sum_{k\in T}v_k\ge t\Big\}\le\exp\left(-t\gamma\ln\left(1+\frac{t\gamma n}{d|I|\,|T|}\right)\right).$$

Proof. Denote

$$x^I:=|I|^{-1/2}\sum_{i\in I}e_i.$$

To simplify the notation, let us assume that $J=\{1,\dots,|J|\}$ (we can permute the degree sequence $\mathbf{d}^{out}$ accordingly). Take any matrix $M\in\mathcal{M}_n(\mathbf{d}^{in},\mathbf{d}^{out})$. Note that, by the assumption on the cardinality of J, we have

$$\big|p_\ell^{col}([k],M)-p_{\ell'}^{col}([k],M)\big|-\big|\mathbf{d}_\ell^{in}-\mathbf{d}_{\ell'}^{in}\big|\le\sqrt{d},\quad k<|J|,\ \ \ell,\ell'\le n.$$

Hence, for any $j\le n$, $k\le|J|$ we have

$$\mathcal{P}_j^{col}([k],M)\le n\sqrt{d}+\sum_{\ell=1}^n|\mathbf{d}_\ell^{in}-d|+n|\mathbf{d}_j^{in}-d|.$$

Note that $\|\cdot\|_1\le n\,\|\cdot\|_{\psi,n}$ by convexity of $\exp(\cdot)$. Then, in view of the assumptions on $\big\|(\mathbf{d}_i^{in}-d)_{i=1}^n\big\|_{\psi,n}$, we get

$$\mathcal{P}_j^{col}([k],M)\le(1+L)n\sqrt{d}+n|\mathbf{d}_j^{in}-d|.$$

Thus, by the triangle inequality,

$$\|\mathcal{P}^{col}([k],M)\|_{\psi,n}\le(1+L)n\sqrt{d}\|(1,1,\dots,1)\|_{\psi,n}+n\big\|(\mathbf{d}_i^{in}-d)_{i=1}^n\big\|_{\psi,n}\le(L+2)n\sqrt{d}$$

for any $k\le|J|$ and $M\in\mathcal{M}_n(\mathbf{d}^{in},\mathbf{d}^{out})$. For every $k\le|J|$, we denote by η_k the random variable

$$\eta_k:=\big|\langle\mathrm{row}_k(\mathbf{M}),x^I\rangle-\mathbb{E}\big[\langle\mathrm{row}_k(\cdot),x^I\rangle\,|\,\mathrm{row}_j(\mathbf{M}),\,j\le k-1\big]\big|.$$

In view of the above estimate of $\|\mathcal{P}^{col}([k],M)\|_{\psi,n}$ and Theorem 3.12, there is $\gamma'(L)\in(0,1)$ such that for any $\lambda>0$ we have

$$\mathbb{E}\Big[e^{\gamma'\lambda\,\sqrt{|I|}\eta_k}\mid \mathrm{row}_j(\mathbf{M}),\, j\le k-1\Big]\le 2\exp\Big(\frac{d|I|}{n}\,g(\lambda)\Big),$$

for every $k\le |J|$ (recall that $\|x^I\|_\infty=|I|^{-1/2}$). Further, for any $k\le|J|$ we have

$$\frac{1}{\sqrt{|I|}}v_k=\frac{1}{\sqrt{|I|}}\big|p_k^{row}(I,\mathbf{M})-\frac{\mathbf{d}_k^{out}|I^c|}{n}\big|=\big|\langle \mathrm{row}_k(\mathbf{M}),x^I\rangle-\frac{\mathbf{d}_k^{out}\sqrt{|I|}}{n}\big|.$$

Thus, using Proposition 3.13 and Lemma 2.3, we get

$$\begin{aligned}v_k&\le\sqrt{|I|}\eta_k+\sqrt{|I|}\;\big|\mathbb{E}\big[\langle \mathrm{row}_k(\cdot),x^I\rangle\mid \mathrm{row}_j(\mathbf{M}),\, j\le k-1\big]-\frac{\mathbf{d}_k^{out}\sqrt{|I|}}{n}\big|\\&\le\sqrt{|I|}\eta_k+\frac{\mu\sqrt{d}|I|}{n}\;\ln\Big(\frac{2n}{|I|}\Big)\end{aligned}$$

for some $\mu=\mu(L)\ge1$. Hence, for any $k\le|J|$ and any $\lambda>0$ we have

$$\mathbb{E}\Big[e^{\gamma'\lambda\,v_k}\mid \mathrm{row}_j(\mathbf{M}),\, j\le k-1\Big]\le 2\exp\Big(\frac{d|I|}{n}g(\lambda)+\gamma'\lambda\frac{\mu\sqrt{d}|I|}{n}\;\ln\Big(\frac{2n}{|I|}\Big)\Big).$$

By Lemma 2.6, this implies that for any subset $T\subset J$ and any $\lambda>0$ we have

$$\mathbb{E}\,e^{\gamma'\lambda\,\sum_{k\in T}v_k}\le 2^{|T|}\exp\Big[|T|\Big(\frac{d|I|}{n}g(\lambda)+\gamma'\lambda\frac{\mu\sqrt{d}|I|}{n}\;\ln\Big(\frac{2n}{|I|}\Big)\Big)\Big].$$

Now, fix any $t\ge4\mu\sqrt{d}|T|$. By the above estimate for the moment generation function and Markov's inequality, we get

$$\begin{aligned}\mathbb{P}\Big\{\sum_{k\in T}v_k\ge t\Big\}&\le\exp\Big[-\gamma'\lambda t+|T|+\frac{d|I||T|}{n}g(\lambda)+\gamma'\lambda|T|\frac{\mu\sqrt{d}|I|}{n}\;\ln\Big(\frac{2n}{|I|}\Big)\Big]\\&\le\exp\Big[-\frac12\gamma'\lambda t+|T|+\frac{d|I||T|}{n}g(\lambda)\Big]\end{aligned}$$

for any $\lambda>0$. It is easy to see that the last espression is minimized for $\lambda:=\ln\big(1+\frac{\gamma' nt}{2d|I||T|}\big)$. Plugging in the value of λ into the exponent, we get

$$\begin{aligned}\mathbb{P}\Big\{\sum_{k\in T}v_k\ge t\Big\}&\le\exp\Big[|T|+\frac12\gamma' t-\frac12\gamma'\lambda t-\frac{d|I||T|}{n}\lambda\Big]\\&=\exp\Big[|T|+\frac{d|I||T|}{n}\Big(\frac{\gamma' nt}{2d|I||T|}-\frac{\gamma' nt}{2d|I||T|}\lambda-\lambda\Big)\Big]\\&=\exp\Big[|T|-\frac{d|I||T|}{n}H\Big(\frac{\gamma' nt}{2d|I||T|}\Big)\Big],\end{aligned}$$

where the function H is defined by (3). Finally, applying the relation (11), we get that for a large enough $K = K(L)$ and all $t \geq K\sqrt{d}|T|$ we have

$$|T| - \frac{d|I||T|}{n} H\Big(\frac{\gamma' n t}{2d|I||T|}\Big) \leq -\frac{d|I||T|}{2n} H\Big(\frac{\gamma' n t}{2d|I||T|}\Big).$$

The result follows. $\square$

Lemma 4.3. *Let $a \in (0,1)$ and suppose that $n^a \leq d$. Further, let the degree sequences $\mathbf{d}^{in}$ and $\mathbf{d}^{out}$, the subset J, the random vectors $v(I) \in \mathbb{R}^J$ and the parameters L and $\gamma(L), K(L)$ be the same as in Lemma 4.2. Then for a sufficiently large universal constant $C_{4.3}$ we have*

$$\mathbb{P}\Big\{ \|v(I)\|_{\psi,|J|} \leq \frac{C_{4.3} K\sqrt{d}}{\gamma a} \text{ for any interval subset } I \subset [n] \text{ of cardinality at most } c_0 n \Big\} \geq 1 - \frac{1}{n}.$$

Proof. Let $C_{4.3}$ be a sufficiently large constant (its value can be recovered from the proof below). Further, let $I \subset [n]$ be a fixed interval subset of $[n]$ of size at most $c_0 n$. In view of Lemma 2.4, for any vector $x \in \mathbb{R}^J$ with $\|x\|_{\psi,|J|} \geq C_{4.3}\gamma^{-1}a^{-1}\sqrt{d}$ there is a natural $t \leq 2\ln(e|J|)$ such that

$$\Big|\Big\{ i \in J : |x_i| \geq \frac{C_{4.3} t\sqrt{d}}{2\gamma a} \Big\}\Big| \geq |J|(2e)^{-t}.$$

In particular, we can write

$$p := \mathbb{P}\Big\{ \|v(I)\|_{\psi,|J|} \geq \frac{C_{4.3}\sqrt{d}}{\gamma a} \Big\} \leq \sum_{t=1}^{\lfloor 2\ln(e|J|)\rfloor} \sum_{\substack{T\subset J,\\ |T| = \lceil |J|\,(2e)^{-t}\rceil}} \mathbb{P}\Big\{ \forall k \in T,\, v_k(I) \geq \frac{C_{4.3} t\sqrt{d}}{2\gamma a} \Big\}.$$

Then, applying Lemma 4.2, we get

$$\begin{aligned} p &\leq \sum_{t=1}^{\lfloor 2\ln(e|J|)\rfloor} \sum_{\substack{T\subset J,\\ |T| = \lceil |J|\,(2e)^{-t}\rceil}} \exp\left[-\frac{C_{4.3} t\sqrt{d}|T|}{2a} \ln\left(1 + \frac{C_{4.3}\, n\, t}{2a\sqrt{d}|I|}\right)\right] \\ &\leq \sum_{t=1}^{\lfloor 2\ln(e|J|)\rfloor} \exp\left[4\lceil |J|(2e)^{-t}\rceil\, t - \frac{C_{4.3} t\sqrt{d}\lceil |J|(2e)^{-t}\rceil}{2a} \ln\left(1 + \frac{C_{4.3}\, n\, t}{2a\sqrt{d}|I|}\right)\right]. \end{aligned}$$

Now using that $\ln\left(1 + \frac{C_{4.3}\, n\, t}{2a\sqrt{d}|I|}\right) \gg \frac{t}{\sqrt{d}} \geq \frac{1}{\sqrt{d}}$ for any t in the above sum, we get

$$\begin{aligned} p &\leq \sum_{t=1}^{\lfloor 2\ln(e|J|)\rfloor} \exp\left(-\frac{C_{4.3} t\lceil |J|(2e)^{-t}\rceil}{4a}\right) \\ &\leq \lfloor 2\ln(e|J|)\rfloor \max_{t=1,\dots,\lfloor 2\ln(e|J|)\rfloor} \exp\left(-\frac{C_{4.3} t|J|(2e)^{-t}}{4a}\right) \ll \frac{1}{n^3}, \end{aligned}$$

where the last inequality follows from the lower bound on d and the choice of $C_{4.3}$. It remains to apply the union bound over all interval subsets (of which there are $O(n^2)$) to finish the proof. □

Remark 4.4. It is easy to see from the proof that the probability estimate $1 - n^{-1}$ in the lemma can be replaced with $1 - n^{-m}$ for any $m > 0$ at the expense of replacing $C_{4.3}$ by a larger constant.

As a consequence of the above, we obtain

Proposition 4.5. *For any parameters $a \in (0,1)$, $L \geq 1$ and $m \in \mathbb{N}$ there is $n_0 = n_0(a,m,L)$ and $\widetilde{L} = \widetilde{L}(L,m)$ (i.e. $\widetilde{L}$ depends only on L and m) with the following property: Let $n \geq n_0$, $n^a \leq d$ and let the degree sequences $\mathbf{d}^{in}$ and $\mathbf{d}^{out}$ be such that $\big\|(\mathbf{d}^{in}_i - d)_{i=1}^n\big\|_{\psi,n}, \big\|(\mathbf{d}^{out}_i - d)_{i=1}^n\big\|_{\psi,n} \leq L\sqrt{d}$. Then the event $\mathcal{E}_{\mathcal{P}}(a^{-1}\widetilde{L})$ (defined by formula (4)) has probability at least $1 - n^{-m}$.*

Proof. Let us partition $[n]$ into at most $2\sqrt{d}$ subsets $J_1, J_2, \ldots, J_r$ ($r \leq 2\sqrt{d}$), where each J_j satisfies $\sqrt{d}/2 \leq |J_j| \leq \sqrt{d}$. For any $j \leq r$, in view of Lemma 4.3, with probability at least $1 - n^{-m-2}$ the $|J_j|$-dimensional vector

$$v^j(I) = (v^j_k)_{k\in J_j}, \quad v^j_k := \Big|p^{row}_k(I,\mathbf{M}) - \frac{\mathbf{d}^{out}_k|I^c|}{n}\Big|, \; k \in J_j,$$

satisfies $\|v^j(I)\|_{\psi,|J^j|} \leq K'a^{-1}\sqrt{d}$ for some $K' = K'(m,L) \geq 1$ for any interval subset $I \subset [n]$ of cardinality at most $c_0 n$. Hence, with probability at least $1 - n^{-m-1}$, the concatenated n-dimensional vector

$$v(I) = (v_1, v_2, \ldots, v_n), \;\; v_k = v^j_k \quad \text{for any } j \leq r \text{ and } k \in J_j$$

satisfies $\|v(I)\|_{\psi,n} \leq K'a^{-1}\sqrt{d}$ for any interval subset $I \subset [n]$ of cardinality at most $c_0 n$. Next, note that for any $k \leq n$ and any $I \subset [n]$ we have

$$\begin{aligned}
nv_k &= n\Big|p^{row}_k(I,\mathbf{M}) - \frac{\mathbf{d}^{out}_k|I^c|}{n}\Big| \\
&\geq \sum_{i=1}^n \big|p^{row}_k(I,\mathbf{M}) - p^{row}_i(I,\mathbf{M})\big| - \sum_{i=1}^n \Big|p^{row}_i(I,\mathbf{M}) - \frac{\mathbf{d}^{out}_i|I^c|}{n}\Big| \\
&\quad - \sum_{i=1}^n \Big|\frac{\mathbf{d}^{out}_i|I^c|}{n} - \frac{d|I^c|}{n}\Big| - |I^c|\big|\mathbf{d}^{out}_k - d\big| \\
&\geq \mathcal{P}^{row}_k(I,\mathbf{M}) - \sum_{i=1}^n v_i - \sum_{i=1}^n \big|\mathbf{d}^{out}_i - d\big| - n\big|\mathbf{d}^{out}_k - d\big|.
\end{aligned}$$

Hence, in view of the convexity of $\exp(\cdot)$, we get

$$\mathcal{P}^{row}_k(I,\mathbf{M}) \leq nv_k + n\big|\mathbf{d}^{out}_k - d\big| + n\big\|(\mathbf{d}^{out}_i - d)_{i=1}^n\big\|_{\psi,n} + n\|v(I)\|_{\psi,n},$$

which implies that

$$\|\mathcal{P}^{row}(I,\mathbf{M})\|_{\psi,n} \leq 2n\big\|(\mathbf{d}^{out}_i - d)_{i=1}^n\big\|_{\psi,n} + 2n\|v(I)\|_{\psi,n}.$$

Therefore, with probability at least $1 - n^{-m-1}$, we have $\|\mathcal{P}^{row}(I,\mathbf{M})\|_{\psi,n} \le \widetilde{L}a^{-1}n\sqrt{d}$ for any interval subset $I \subset [n]$ of cardinality at most $c_0 n$ and $\widetilde{L} := 2K' + 2L$. Clearly, the same estimate holds for $\mathcal{P}^{col}(I,\mathbf{M})$ and the proof is complete. □

4.2. Concentration inequality for linear forms

The goal of this subsection is to provide the proof of Theorem D. Let us introduce a family of random variables on the probability space $(\mathcal{M}_n(\mathbf{d}^{in},\mathbf{d}^{out}),\mathbb{P})$ as follows. Take any index $i \le n$ and any subset $I \subset [n]$ not containing i. Further, let $x \in \mathbb{R}^n$ be any vector. Then we define $\theta(i,I,x) : \mathcal{M}_n(\mathbf{d}^{in},\mathbf{d}^{out}) \to \mathbb{R}$ as

$$\theta(i,I,x) := \mathbb{E}\big[\langle \mathrm{row}_i(\mathbf{M}), x\rangle \mid \mathrm{row}_j(\mathbf{M}),\ j \in I\big].$$

In other words, $\theta(i,I,x)$ is the conditional expectation of $\langle \mathrm{row}_i(\mathbf{M}), x\rangle$, conditioned on realizations of rows $\mathrm{row}_j(\mathbf{M})$ $(j \in I)$.

Lemma 4.6. *Let $L > 0$ be some parameter and let the event $\mathcal{E}_{\mathcal{P}}(L)$ be defined by (4). Let I be any non-empty interval subset of $[n]$ of length at most $c_0 n$ and let $Q = (Q_{ij})$ be a fixed $n \times n$ matrix with all entries with indices outside $I \times [n]$ equal to zero. Then for any $t > 0$ we have*

$$\mathbb{P}\Big\{\Big|\sum_{i\in I}\Big(\sum_{j=1}^{n}\mathbf{M}_{ij}Q_{ij} - \theta\big(i,\{\inf I,\dots,i-1\},\mathrm{row}_i(Q)\big)\Big)\Big| > t \mid \mathbf{M} \in \mathcal{E}_{\mathcal{P}}(L)\Big\}$$
$$\le \frac{2}{\mathbb{P}(\mathcal{E}_{\mathcal{P}}(L))}\exp\left(-\frac{d\,\|Q\|_{HS}^2}{n\,\|Q\|_\infty^2}\,H\left(\frac{\gamma t n\|Q\|_\infty}{d\|Q\|_{HS}^2}\right)\right),$$

where $\gamma = \gamma(L)$ is taken from Theorem 3.12.

Proof. Fix for a moment any $i \in I$ and let

$$\mathcal{E}_i := \big\{\|\mathcal{P}^{col}(\{\inf I,\dots,i-1\},\mathbf{M})\|_{\psi,n} \le Ln\sqrt{d}\big\}.$$

Further, denote by η_i the random variable

$$\eta_i := \left[\sum_{j=1}^{n}\mathbf{M}_{ij}Q_{ij} - \theta\big(i,\{\inf I,\dots,i-1\},\mathrm{row}_i(Q)\big)\right]\chi_i,$$

where χ_i is the indicator function of the event $\mathbf{M} \in \mathcal{E}_i$. Note that $\|\mathcal{P}^{col}(\{\inf I,\dots,i-1\},\mathbf{M})\|_{\psi,n}$ is uniquely determined by realizations of $\mathrm{row}_{\inf I}(\mathbf{M}),\dots,\mathrm{row}_{i-1}(\mathbf{M})$. Now, assume that Y_j $(j = \inf I,\dots,i-1)$ is any realization of rows $\mathrm{row}_j(\mathbf{M})$ $(j = \inf I,\dots,i-1)$ such that, conditioned on this realization, $\mathbf{M}$ belongs to $\mathcal{E}_i$. That is,

$$\big\{\mathrm{row}_j(\mathbf{M}) = Y_j,\ \ j = \inf I,\dots,i-1\big\} \subset \mathcal{E}_i.$$

Then, applying Theorem 3.12, we obtain

$$\mathbb{E}\big[e^{\gamma\lambda\sum_{j=1}^{n}\mathbf{M}_{ij}Q_{ij} - \gamma\lambda\theta(i,\{\inf I,\dots,i-1\},\mathrm{row}_i(Q))} \mid \mathrm{row}_j(\mathbf{M}) = Y_j,\ j = \inf I,\dots,i-1\big]$$
$$\le \exp\left(\frac{d\,\|\mathrm{row}_i(Q)\|^2}{n\max_{j\le n}{Q_{ij}}^2}\,g(\lambda\max_{j\le n}Q_{ij})\right),\quad \lambda > 0,$$

for some $\gamma = \gamma(L)$. Note that the value of η_j is uniquely determined by realizations of rows $\mathrm{row}_k(\mathbf{M})$ $(k \le j)$. Hence, in view of the definition of η_i, we get from the last relation

$$\mathbb{E}\big[e^{\lambda\eta_i} \mid \eta_j,\ j = \inf I, \dots, i-1\big] \le \exp\left(\frac{d\,\|\mathrm{row}_i(Q)\|^2}{n \max_{j\le n} {Q_{ij}}^2}\, g(\lambda\gamma^{-1}\max_{j\le n} Q_{ij})\right),\ \ \lambda > 0.$$

Now, let

$$\eta := \sum_{i\in I} \eta_i.$$

By the above inequality and by Corollary 2.6, we get

$$\mathbb{P}\big\{\eta \ge t\big\} \le \exp\left(-\frac{d\,\|Q\|_{HS}^2}{n\,\|Q\|_\infty^2}\, H\left(\frac{\gamma t n \|Q\|_\infty}{d\|Q\|_{HS}^2}\right)\right), \quad t > 0.$$

Finally, note that

$$\mathcal{E}_{\mathcal{P}}(L) \subset \bigcap_{i\in I} \mathcal{E}_i,$$

whence, restricted to $\mathcal{E}_{\mathcal{P}}(L)$, the variable η is equal to

$$\sum_{i\in I}\Big(\sum_{j=1}^n \mathbf{M}_{ij}Q_{ij} - \theta\big(i, \{\inf I, \dots, i-1\}, \mathrm{row}_i(Q)\big)\Big).$$

It follows that

$$\mathbb{P}\Big\{\sum_{i\in I}\Big(\sum_{j=1}^n \mathbf{M}_{ij}Q_{ij} - \theta\big(i, \{\inf I, \dots, i-1\}, \mathrm{row}_i(Q)\big)\Big) \ge t \mid \mathbf{M} \in \mathcal{E}_{\mathcal{P}}(L)\Big\}$$
$$\le \frac{1}{\mathbb{P}(\mathcal{E}_{\mathcal{P}}(L))} \exp\left(-\frac{d\,\|Q\|_{HS}^2}{n\,\|Q\|_\infty^2}\, H\left(\frac{\gamma t n\|Q\|_\infty}{d\|Q\|_{HS}^2}\right)\right), \quad t > 0.$$

Applying a similar argument to the variable $-\eta$, we get the result. □

The next lemma allow us to replace the variables $\theta\big(i, \{\inf I, \dots, i-1\}, \mathrm{row}_i(Q)\big)$ with constants.

Lemma 4.7. *For any $L \ge 1$ there is $n_0 = n_0(L)$ with the following property. Let $n \ge n_0$, let I be any non-empty interval subset of $[n]$ of length at most $c_0 n$ and let $Q = (Q_{ij})$ be a fixed $n \times n$ matrix with all entries with indices outside $I \times [n]$ equal to zero. Then*

$$\Big|\sum_{i\in I}\theta\big(i, \{\inf I, \dots, i-1\}, \mathrm{row}_i(Q)\big) - \sum_{i\in I}\frac{\mathbf{d}_i^{out}}{n}\sum_{j=1}^n Q_{ij}\Big| \le C_{4.7} L\sqrt{d}\sum_{i\in I}\|\mathrm{row}_i(Q)\|_{\log,n}$$

everywhere on $\mathcal{E}_{\mathcal{P}}(L)$. Here, $C_{4.7} > 0$ is a universal constant.

Proof. In view of the relation $\|\cdot\|_1 \le en\|\cdot\|_{\log,n}$ which follows from convexity of the function $t\ln_+(t)$, it is enough to show that for any $i\in I$ we have

$$\Big|\theta\big(i,\{\inf I,\dots,i-1\},\mathrm{row}_i(Q)\big)-\frac{\mathbf{d}_i^{out}}{n}\sum_{j=1}^n Q_{ij}\Big|$$
$$\le \frac{C\sqrt{d}}{n}\|\mathrm{row}_i(Q)\|_1+\frac{C}{n}\|\mathrm{row}_i(Q)\|_{\log,n}\|\mathcal{P}^{col}(\{\inf I,\dots,i-1\},\mathbf{M})\|_{\psi,n}$$

everywhere on $\mathcal{E}_{\mathcal{P}}(L)$ for a sufficiently large constant $C>0$. But this follows immediately from Proposition 3.13. □

Finally, we can prove the main technical result of the paper. To make the statement self-contained, we explicitly mention all the assumptions on parameters. Given an $n\times n$ matrix Q, we define *the shift* $\Delta(Q)$ as

$$\Delta(Q):=\sqrt{d}\sum_{i=1}^n\|\mathrm{row}_i(Q)\|_{\log,n}.$$

Theorem 4.8. *For any $L\ge 1$ there are $\gamma=\gamma(L)>0$ and $n_0=n_0(L)$ with the following properties. Assume that $n\ge n_0$ and that the degree sequences $\mathbf{d}^{in},\mathbf{d}^{out}$ satisfy*

$$(1-c_0)d\le \mathbf{d}_i^{in},\mathbf{d}_i^{out}\le d,\quad i\le n$$

for some natural d with $C_1\ln^2 n\le d\le(1/2+c_0)n$. Further, assume that the set $\mathcal{M}_n(\mathbf{d}^{in},\mathbf{d}^{out})$ is non-empty. Then, with $\mathcal{E}_{\mathcal{P}}(L)$ defined by (4), we have for any $n\times n$ matrix Q:

$$\mathbb{P}\Big\{\Big|\sum_{i=1}^n\sum_{j=1}^n\mathbf{M}_{ij}Q_{ij}-\sum_{i=1}^n\frac{\mathbf{d}_i^{out}}{n}\sum_{j=1}^n Q_{ij}\Big|>t+C_2L\Delta(Q)\ \mid\ \mathbf{M}\in\mathcal{E}_{\mathcal{P}}(L)\Big\}$$
$$\le\frac{C_3}{\mathbb{P}(\mathcal{E}_{\mathcal{P}}(L))}\exp\left(-\frac{d\,\|Q\|_{HS}^2}{n\,\|Q\|_\infty^2}\,H\left(\frac{\gamma tn\|Q\|_\infty}{d\|Q\|_{HS}^2}\right)\right),\quad t>0.$$

Here, $C_1,C_2,C_3>0$ are sufficiently large universal constants.

Proof. Let us partition $[n]$ into $\lceil 2/c_0\rceil$ interval subsets I_j ($j\le\lceil 2/c_0\rceil$), with each I_j of cardinality at most c_0n. Further, define $n\times n$ matrices Q^j ($j\le\lceil 2/c_0\rceil$) as

$$Q^j_{k,\ell}:=\begin{cases}Q_{k\ell}, & \text{if } k\in I^j;\\ 0, & \text{otherwise.}\end{cases}$$

Note that each Q^j satisfies assumptions of both Lemma 4.6 and Lemma 4.7. Combining the lemmas, we get

$$\mathbb{P}\Big\{\Big|\sum_{k\in I^j}\sum_{\ell=1}^n\mathbf{M}_{k\ell}Q_{k\ell}-\sum_{k\in I^j}\frac{\mathbf{d}_k^{out}}{n}\sum_{\ell=1}^n Q_{k\ell}\Big|>t+CL\Delta_j\ \mid\ \mathbf{M}\in\mathcal{E}_{\mathcal{P}}(L)\Big\}$$
$$\le\frac{2}{\mathbb{P}(\mathcal{E}_{\mathcal{P}}(L))}\exp\left(-\frac{d\,\|Q^j\|_{HS}^2}{n\,\|Q^j\|_\infty^2}\,H\left(\frac{\gamma tn\|Q^j\|_\infty}{d\|Q^j\|_{HS}^2}\right)\right),\quad t>0,$$

where $\Delta_j := \sqrt{d}\sum_{k\in I^j}\|\mathrm{row}_k(Q)\|_{\log,n}$. It is not difficult to check that the function $f(s,w) := \frac{s^2}{w^2}H(\frac{bw}{s^2})$ is decreasing in both arguments s and w for any value of parameter $b>0$. Hence, the above quantity is majorized by

$$\frac{2}{\mathbb{P}(\mathcal{E}_{\mathcal{P}}(L))}\exp\left(-\frac{d\,\|Q\|_{HS}^2}{n\,\|Q\|_\infty^2}\,H\left(\frac{\gamma tn\|Q\|_\infty}{d\|Q\|_{HS}^2}\right)\right).$$

Finally, note that if for some matrix $M\in\mathcal{M}_n(\mathbf{d}^{in},\mathbf{d}^{out})$ and $t>0$ we have

$$\Big|\sum_{k=1}^n\sum_{\ell=1}^n M_{k\ell}Q_{k\ell}-\sum_{k=1}^n\frac{\mathbf{d}_k^{out}}{n}\sum_{\ell=1}^n Q_{k\ell}\Big|>t+CL\Delta(Q)$$

then necessarily

$$\Big|\sum_{k\in I^j}\sum_{\ell=1}^n M_{k\ell}Q_{k\ell}-\sum_{k\in I^j}\frac{\mathbf{d}_k^{out}}{n}\sum_{\ell=1}^n Q_{k\ell}\Big|>\frac{t}{\lceil 2/c_0\rceil}+CL\Delta_j$$

for some $j\le\lceil 2/c_0\rceil$. The result follows. □

Remark 4.9. It is easy to see that constant C_3 in the above theorem can be replaced by any number strictly greater than one, at the expense of decreasing γ.

Remark 4.10. Note that, in view of Lemma 2.3, we have

$$\Delta(Q)\le C_{2.3}\sqrt{\frac{d}{n}}\sum_{i=1}^n\|\mathrm{row}_i(Q)\|\le C_{2.3}\sqrt{d}\|Q\|_{HS}.$$

In particular, if x and y are unit vectors in $\mathbb{R}^n$ then $\Delta(xy^T)\le C_{2.3}\sqrt{d}$. Further, if all non-zero entries of the matrix Q are located in a submatrix of size $k\times\ell$ (for some $k,\ell\le n$) then, again applying Lemma 2.3, we get

$$\Delta(Q)\le C_{2.3}\frac{\sqrt{d\ell}}{n}\ln\frac{2n}{\ell}\,\sum_{i=1}^n\|\mathrm{row}_i(Q)\|\le C_{2.3}\frac{\sqrt{dk\ell}}{n}\ln\frac{2n}{\ell}\,\|Q\|_{HS}.$$

In particular, given a k-sparse unit vector x and an ℓ-sparse unit vector y, we have

$$\Delta(xy^T)\le C_{2.3}\frac{\sqrt{dk\ell}}{n}\ln\frac{2n}{\ell}.$$

Remark 4.11. Assume that $\big\|(\mathbf{d}_i^{out}-d)_{i=1}^n\big\|_{\psi,n}\le K\sqrt{d}$ for some parameter $K>0$. Then we have, in view of Lemma 2.2:

$$\begin{aligned}\Big|\sum_{i=1}^n\frac{\mathbf{d}_i^{out}}{n}\sum_{j=1}^n Q_{ij}-\frac{d}{n}\sum_{i,j=1}^n Q_{ij}\Big| &\le\frac{1}{n}\sum_{j=1}^n\Big|\sum_{i=1}^n(\mathbf{d}_i^{out}-d)Q_{ij}\Big|\\ &\le\frac{1}{n}\big\|(\mathbf{d}_i^{out}-d)_{i=1}^n\big\|\sum_{j=1}^n\|\mathrm{col}_j(Q)\|\\ &\le C_{2.2}K\sqrt{d}\|Q\|_{HS}.\end{aligned}$$

Together with Remark 4.10, this implies that the quantity $\sum_{i=1}^n \frac{\mathbf{d}_i^{out}}{n} \sum_{j=1}^n Q_{ij}$ in the estimate of Theorem 4.8 can be replaced with $\frac{d}{n} \sum_{i,j=1}^n Q_{ij}$ at the expense of substituting $\sqrt{d}\|Q\|_{HS}$ for the shift $\Delta(Q)$.

Remark 4.12. The Bennett–type concentation inequality for linear forms obtained in [12] (see formula (6) there) contains a parameter playing the same role as shift $\Delta(Q)$ in our theorem. However, the dependence of this parameter in [12] on the matrix Q is fundamentally different from ours. Given a random matrix $\widetilde{\mathbf{M}}$ uniformly distributed on $\mathcal{S}_n(d)$, the set of adjacency matrices of undirected simple d-regular graphs on $[n]$, for every matrix Q with non-negative entries and zero diagonal, Theorem 5.1 of [12] gives:

$$\mathbb{P}\Big\{\Big|\sum_{i,j=1}^n \widetilde{\mathbf{M}}_{ij}Q_{ij} - \frac{d}{n}\sum_{i,j=1}^n Q_{ij}\Big| \geq t + \frac{Cd^2}{n^2}\sum_{i,j=1}^n Q_{ij}\Big\} \leq 2\exp\left(-\frac{d\,\|Q\|_{HS}^2}{n\,\|Q\|_\infty^2}\,H\left(\frac{ctn\|Q\|_\infty}{d\|Q\|_{HS}^2}\right)\right).$$

In view of (8), the shift $\frac{d^2}{n^2}\sum_{i,j=1}^n Q_{ij}$ is majorized by $\frac{ed^2}{n}\sum_{i=1}^n \|\mathrm{row}_i(Q)\|_{\log,n} = \frac{ed^{3/2}}{n}\Delta(Q)$. Thus, the concentration inequality from [12] gives sharper estimates than ours provided that $d = O(n^{2/3})$. On the other hand, for $d \gg n^{2/3}$ the estimate in [12] becomes insufficient to produce the optimal upper bound on the matrix norm, whereas our shift $\Delta(Q)$ gives satisfactory estimates for all large enough d. Let us emphasize that this comparison is somewhat artificial since [12] deals only with undirected graphs and symmetric matrices, while our Theorem 4.8 applies to the directed setting.

The proof of Theorem D from the Introduction is obtained by combining Theorem 4.8 with Remarks 4.9–4.11 and Proposition 4.5.

Let us finish this section by discussing the necessity of the tensorization procedure. As we mentioned in the Introduction, Freedman's inequality for martingales was employed in paper [13] dealing with *the permutation model* of regular graphs (when the adjacency matrix of corresponding random multigraph is constructed using independent random permutation matrices and their transposes). It was proved in [13] that the second largest eigenvalue of such a graph is of order $O(\sqrt{d})$ with high probability. Importantly, in [13] the martingale sequence was constructed for *the entire* matrix, thereby yielding a concentration inequality directly after applying Freedman's theorem and without any need for a tensorization procedure. The fact that in our paper we construct martingales row by row is essentially responsible for the presence of the "shift" $\Delta(Q)$ in our concentration inequality, and forced us to develop the lengthy and technical tensorization. However, when constructing a single martingale sequence over the entire matrix, revealing the matrix entries one by one in some appropriate order, it is not clear to us how to control martingale's parameters (absolute values of the differences and their variances). Nevertheless, it seems natural to expect that some kind of an "all-matrix" martingale can be constructed and analysed, yielding a much stronger concentration inequality for linear forms.

5. The Kahn–Szeméredi argument

In this section, we use the concentration result established above and the well known argument of Kahn and Szeméredi [18] to bound $s_2(\mathbf{M})$, for $\mathbf{M}$ uniformly distributed on $\mathcal{M}_n(\mathbf{d}^{in}, \mathbf{d}^{out})$. We refer to [12] for a detailed exposition of Kahn-Szeméredi's argument in the setting of undirected graphs. The adaptation to the directed setting is straightforward (and essentially notational) and we leave it to the interested reader. Set

$$S_0^{n-1} := \Big\{y \in S^{n-1} : \sum_{i=1}^n y_i = 0\Big\}.$$

The Courant–Fischer formula implies

$$s_2(\mathbf{M}) \leq \sup_{y\in S_0^{n-1}} \|\mathbf{M}y\| = \sup_{(x,y)\in S^{n-1}\times S_0^{n-1}} \langle \mathbf{M}y, x\rangle$$

(of course, the above relation is true for *any* $n \times n$ matrix M). To estimate the expression on the right hand side, we shall apply our concentration inequality to $\langle \mathbf{M}y, x\rangle$ for any fixed couple $(x, y) \in S^{n-1} \times S_0^{n-1}$, and then invoke a covering argument. Let us take a closer look at the procedure. We have for any admissible x, y:

$$\langle \mathbf{M}y, x\rangle = \sum_{i,j=1}^n x_i \mathbf{M}_{ij} y_j = \sum_{i,j=1}^n \mathbf{M}_{ij} Q_{ij},$$

where $Q := xy^t$ satisfies $\|Q\|_{\mathrm{HS}} = 1$ and $\|Q\|_\infty = \max\limits_{i,j\in[n]} |x_i y_j| = \|x\|_\infty \|y\|_\infty$. Therefore, in view of the concentration statement obtained in Section 4, the (conditional) probability that $\langle \mathbf{M}y, x\rangle \gg \sqrt{d}$ is bounded by

$$\exp\bigg(-\frac{d}{n}\frac{H\big(\frac{n}{\sqrt{d}}\|x\|_\infty \|y\|_\infty\big)}{\|x\|_\infty^2 \|y\|_\infty^2}\bigg)$$

(we disregard any constant factors in the above expression). However, when $\|x\|_\infty \|y\|_\infty \gg \sqrt{d}/n$, the estimate becomes too weak (larger than C^{-n}) to apply the union bound over a net of size exponential in n. The idea of Kahn and Szeméredi is to split the entries of Q into two groups according to their magnitude. Then the standard approach discussed above would work for the collection of entries smaller than $\sqrt{d}/n$. Corresponding pairs of indices are called *light couples*. For the second group, the key idea is to exploit discrepancy properties of the associated graph; again, our concentration inequality will play a crucial role in their verification.

Given $(x, y) \in S^{n-1} \times S_0^{n-1}$, let us define

$$\mathcal{L}(x,y) := \big\{(i,j) \in [n]^2 : |x_i y_j| \leq \sqrt{d}/n\big\} \quad \text{and} \quad \mathcal{H}(x,y) := \big\{(i,j) \in [n]^2 : |x_i y_j| > \sqrt{d}/n\big\}.$$

The notation $\mathcal{L}(x, y)$ stands for *light couples* while $\mathcal{H}(x, y)$ refers to *heavy couples*. Moreover, we will represent the corresponding partition of Q as $Q =$

$Q_{\mathcal{L}} + Q_{\mathcal{H}}$, where $Q_{\mathcal{L}}, Q_{\mathcal{H}}$ are both $n \times n$ matrices in which the entries from "the alien" collection are replaced with zeros.

Throughout the section, we always assume that the degree sequences $\mathbf{d}^{in}$, $\mathbf{d}^{out}$ satisfy (6) for some d, and that d itself satisfies (29). Moreover, we always assume that the set of matrices $\mathcal{M}_n(\mathbf{d}^{in}, \mathbf{d}^{out})$ is non-empty. As before, $\mathbf{M}$ is the random matrix uniformly distributed on $\mathcal{M}_n(\mathbf{d}^{in}, \mathbf{d}^{out})$ and $\mathbf{G}$ is the associated random graph.

Lemma 5.1. *For any $L \geq 1$ there is $\gamma = \gamma(L) > 0$ with the following property: Let $n \geq C_{5.1}$ and let $(x, y) \in S^{n-1} \times S_0^{n-1}$. Then for any $t > 0$ we have*

$$\mathbb{P}\Big\{\Big| \sum_{(i,j)\in\mathcal{L}(x,y)} x_i \mathbf{M}_{ij} y_j\Big| \geq (C_{5.1}\, L + t)\sqrt{d} \;|\; \mathcal{E}_{\mathcal{P}}(L)\Big\} \leq \frac{C_{5.1}}{\mathbb{P}(\mathcal{E}_{\mathcal{P}}(L))} \exp\big(-n\, H(\gamma\, t)\big).$$

Here, $C_{5.1} > 0$ is a sufficiently large universal constant and $\mathcal{E}_{\mathcal{P}}(L)$ is defined by (4).

Proof. Let $(x, y) \in S^{n-1} \times S_0^{n-1}$ and denote $Q := xy^t$. Let $Q_{\mathcal{L}}$ and $Q_{\mathcal{H}}$ be defined as above. By the definition of $\mathcal{L}(x, y)$, we have $\|Q_{\mathcal{L}}\|_\infty \leq \sqrt{d}/n$, and, since $\|x\| = \|y\| = 1$, we have $\|Q_{\mathcal{L}}\|_{\mathrm{HS}} \leq 1$. Further, note that

$$\sum_{i=1}^{n} \|\mathrm{row}_i(Q_{\mathcal{L}})\|_1 \leq \|x\|_1\, \|y\|_1 \leq n,$$

whence, in view of Lemma 2.3,

$$\sum_{i=1}^{n} \|\mathrm{row}_i(Q_{\mathcal{L}})\|_{\log,n} \leq C_{2.3}.$$

Applying Theorem 4.8 to matrix $Q_{\mathcal{L}}$ with $t := r\sqrt{d}$ ($r > 0$), we get that there exists $\gamma := \gamma(L) > 0$ depending on L such that

$$\begin{aligned}\mathbb{P}\Big\{\Big| \sum_{(i,j)\in\mathcal{L}(x,y)} x_i \mathbf{M}_{ij} y_j - \sum_{(i,j)\in\mathcal{L}(x,y)} \frac{\mathbf{d}_i^{out}\, x_i y_j}{n}\Big| \geq (C\, L + r)\sqrt{d} \;|\; \mathbf{M} \in \mathcal{E}_{\mathcal{P}}(L)\Big\} \\ \leq \frac{C_3}{\mathbb{P}(\mathcal{E}_{\mathcal{P}}(L))} \exp\big(-n\, H(\gamma\, r)\big),\end{aligned} \tag{35}$$

where C is a universal constant and C_3 is the constant from Theorem 4.8. Since the coordinates of y sum up to zero, we have for any $i \leq n$:

$$\Big| \sum_{j:\,(i,j)\in\mathcal{L}(x,y)} \mathbf{d}_i^{out}\, x_i y_j\Big| = \Big| \sum_{j:\,(i,j)\in\mathcal{H}(x,y)} \mathbf{d}_i^{out}\, x_i y_j\Big| \leq d \sum_{j:\,(i,j)\in\mathcal{H}(x,y)} \frac{(x_i y_j)^2}{\sqrt{d}/n},$$

where in the last inequality we used that $\mathbf{d}_i^{out} \leq d$ and $|x_i y_j| \geq \sqrt{d}/n$ for $(i, j) \in \mathcal{H}(x, y)$. Summing over all rows and using the condition $\|x\| = \|y\| = 1$, we get

$$\Big| \sum_{(i,j)\in\mathcal{L}(x,y)} \frac{\mathbf{d}_i^{out}\, x_i y_j}{n}\Big| \leq \sqrt{d}.$$

This, together with (35), finishes the proof after choosing $C_{5.1} \geq C + 1$. $\square$

Next, we prove a discrepancy property for our model. In what follows, for any subsets $S,T\subset[n]$, $\mathbf{E}_G(S,T)$ denotes the set of edges of $\mathbf{G}$ emanating from S and landing in T. For any $K_1,K_2\geq 1$, we denote by $\mathcal{E}_{5.2}(K_1,K_2)$ the event that for *all* subsets $S,T\subset[n]$ at least one of the following is true:

$$|\mathbf{E}_G(S,T)|\leq K_1\,\frac{d}{n}\,|S|\,|T|, \tag{36}$$

or

$$|\mathbf{E}_G(S,T)|\,\ln\left(\frac{|\mathbf{E}_G(S,T)|}{\frac{d}{n}\,|S|\,|T|}\right)\leq K_2\max(|S|,|T|)\,\ln\left(\frac{e\,n}{\max(|S|,|T|)}\right). \tag{37}$$

Let us note that both conditions above can be equivalently restated using a single formula; however, the presentation in form (36)–(37) nicely captures the underlying dichotomy within a "typical" realization of $\mathbf{G}$: either both S and T are "large", in which case the edge count does not deviate too much from its expectation, or at least one of the sets is "small", and the edge count, up to a logarithmic multiple, is bounded by the cardinality of the larger vertex set.

Proposition 5.2. *For any $L\geq 1$ and $m\in\mathbb{N}$ there are $n_0=n_0(L,m)$, $K_1=K_1(L,m)$ and $K_2=K_2(L,m)$ such that for $n\geq n_0$ and d satisfying (29) we have*

$$\mathbb{P}\big(\mathcal{E}_{5.2}(K_1,K_2)\mid\mathcal{E}_{\mathcal{P}}(L)\big)\geq 1-\frac{1}{\mathbb{P}(\mathcal{E}_{\mathcal{P}}(L))\,n^m}.$$

Proof. Fix for a moment any $S,T\subset[n]$ and let Q be the $n\times n$ matrix whose entries are equal to 1 on $S\times T$ and 0 elsewhere. Set $k:=|S|$ and $\ell:=|T|$. From Remark 4.10, we have

$$\Delta(Q)\leq C_{2.3}\frac{\sqrt{dk\ell}}{n}\ln\frac{2n}{\ell}\,\|Q\|_{HS}\leq C_{2.3}\frac{\sqrt{d}k\ell}{n}\ln(2n)\leq C_{2.3}\frac{dk\ell}{n},$$

where the last inequality follows from the assumption (29) on d. Using the estimate together with the inequality $\mathbf{d}_i^{out}\leq d$ $(i\leq n)$ and applying Theorem 4.8, for any $r>0$ we obtain

$$\mathbb{P}\Big\{|\mathbf{E}_G(S,T)|>(CL{+}r)\frac{d}{n}k\,\ell\;\mid\;\mathbf{M}\in\mathcal{E}_{\mathcal{P}}(L)\Big\}\leq\frac{C}{\mathbb{P}(\mathcal{E}_{\mathcal{P}}(L))}\,\exp\left(-\frac{d\,k\,\ell}{n}\,H(\gamma\,r)\right), \tag{38}$$

for a universal constant $C\geq 1$ and some $\gamma=\gamma(L)>0$. Now, we set $K_1:=2CL$ and let $K_2=K_2(L,m)$ be the minimum number such that $K_2H(\gamma t)\geq 2(3+m)t\ln(2t)$ for all $t\geq CL$ (note that the definition of K_1, K_2 does not depend on S and T). Since the function H is strictly increasing on $(0,\infty)$, there is a unique number $r_1>0$ such that

$$H(\gamma r_1)=\frac{(3+m)\max(k,\ell)}{\frac{d}{n}\,k\,\ell}\,\ln\left(\frac{e\,n}{\max(k,\ell)}\right).$$

Next, note that if for a fixed realization of the graph $\mathbf{G}$ we have $|\mathbf{E}_G(S,T)|\leq(CL+r_1)\,\frac{d}{n}k\,\ell$ then either (36) or (37) holds. Indeed, if $r_1\leq CL$ then the assertion is obvious. Otherwise, if $r_1 > CL$ then, by the definition of K_2, we have $(CL+r_1)\ln(CL+r_1) \le \frac{K_2}{3+m}\,H(\gamma r_1)$. Together with the trivial estimate

$$|\mathbf{E}_G(S,T)|\,\ln\left(\frac{|\mathbf{E}_G(S,T)|}{\frac{d}{n}\,k\,\ell}\right) \le \frac{d}{n}k\,\ell\,(CL+r_1)\ln(CL+r_1),$$

this gives

$$|\mathbf{E}_G(S,T)|\,\ln\left(\frac{|\mathbf{E}_G(S,T)|}{\frac{d}{n}\,k\,\ell}\right) \le \frac{K_2}{3+m}\,\frac{d}{n}k\,\ell\,H(\gamma r_1) = K_2\max(k,\ell)\,\ln\left(\frac{e\,n}{\max(k,\ell)}\right).$$

Thus, all realizations of $\mathbf{M}$ (or, equivalently, $\mathbf{G}$) with $|\mathbf{E}_G(S,T)| \le (CL + r_1)\,\frac{d}{n}|S|\,|T|$ for all $S,T\subset[n]$, necessarily fall into event $\mathcal{E}_{5.2}(K_1,K_2)$. It follows that

$$\mathbb{P}\big(\mathcal{E}^c_{5.2}(K_1,K_2)\mid\mathcal{E}_{\mathcal{P}}(L)\big) \le \mathbb{P}\Big\{\exists S,T\subset[n]:\,|\mathbf{E}_G(S,T)| > (CL+r_1)\frac{d}{n}|S|\,|T|\;\mid\;\mathbf{M}\in\mathcal{E}_{\mathcal{P}}(L)\Big\}.$$

Applying (38), we get

$$\begin{aligned}
\mathbb{P}\big(\mathcal{E}^c_{5.2}(K_1,K_2)\mid\mathcal{E}_{\mathcal{P}}(L)\big) &\le \frac{C}{\mathbb{P}(\mathcal{E}_{\mathcal{P}}(L))}\,\sum_{k,\ell=1}^{n}\binom{n}{k}\binom{n}{\ell}\exp\left(-\frac{d\,k\ell}{n}\,H(\gamma r_1)\right)\\
&\le \frac{C}{\mathbb{P}(\mathcal{E}_{\mathcal{P}}(L))}\,\sum_{k,\ell=1}^{n}\exp\left(k\ln\big(\frac{en}{k}\big)+\ell\ln\big(\frac{en}{\ell}\big)-\frac{d\,k\ell}{n}\,H(\gamma r_1)\right)\\
&\le \frac{C}{\mathbb{P}(\mathcal{E}_{\mathcal{P}}(L))}\,\sum_{k,\ell=1}^{n}\exp\left[-(m+1)\max(k,\ell)\,\ln\left(\frac{e\,n}{\max(k,\ell)}\right)\right]\\
&\le \frac{C}{\mathbb{P}(\mathcal{E}_{\mathcal{P}}(L))\,n^{m+1}}\\
&\le \frac{1}{\mathbb{P}(\mathcal{E}_{\mathcal{P}}(L))\,n^{m}},
\end{aligned}$$

where we used the estimate $\max(k,\ell)\,\ln\left(\frac{e\,n}{\max(k,\ell)}\right) \ge \ln n$. □

The conditions on the edge count of a graph expressed via (36) or (37), are a basic element in the argument of Kahn and Szeméredi. The following lemma shows that the contribution of heavy couples to the matrix norm is *deterministically* controlled once we suppose that either (36) or (37) holds for all vertex subsets of corresponding graph.

Lemma 5.3. *For any $K_1,K_2>0$ there exists $\beta>0$ depending only on K_1,K_2 such that the following holds. Let, as usual, the degree sequences $\mathbf{d}^{in},\mathbf{d}^{out}$ be bounded from above by d (coordinate-wise) and let $M\in\mathcal{E}_{5.2}(K_1,K_2)$. Then for any $x,y\in S^{n-1}$, we have*

$$\Big|\sum_{(i,j)\in\mathcal{H}(x,y)} x_i M_{ij} y_j\Big| \le \beta\sqrt{d}.$$

A proof of this statement is well known [18, 12, 13]. In [12], it is stated in the undirected setting while in [18, 13] it is formulated for the directed permutation model. Nevertheless, in our case, the proof remains the same and requires no modifications compared to the ones mentioned above.

In order to simultaneously estimate contribution of all pairs of vectors from $S^{n-1} \times S_0^{n-1}$ to the second largest singular value of our random matrix, we shall discretize this set. The following lemma is quite standard.

Lemma 5.4. *Let $\varepsilon \in (0, 1/2)$, $\mathcal{N}_\varepsilon$ be a Euclidean ε-net in S^{n-1}, and $\mathcal{N}_\varepsilon^0$ be a Euclidean ε-net in S_0^{n-1}. Further, let A be any $n \times n$ non-random matrix and R be any positive number such that $|\langle Ax, y\rangle| \le R$ for all $(x, y) \in \mathcal{N}_\varepsilon \times \mathcal{N}_\varepsilon^0$. Then $|\langle Ax, y\rangle| \le R/(1-2\varepsilon)$ for all $(x, y) \in S^{n-1} \times S_0^{n-1}$.*

Proof. Let $(x_0, y_0) \in S^{n-1} \times S_0^{n-1}$ be such that $a := \sup_{(x,y)\in S^{n-1}\times S_0^{n-1}} \langle Ax, y\rangle = \langle Ax_0, y_0\rangle$. By the definition of $\mathcal{N}_\varepsilon$ and $\mathcal{N}_\varepsilon^0$, there exists a pair $(x_0', y_0') \in \mathcal{N}_\varepsilon \times \mathcal{N}_\varepsilon^0$ such that $\|x_0 - x_0'\| \le \varepsilon$ and $\|y_0 - y_0'\| \le \varepsilon$. Together with the fact that the normalized difference of two elements in S_0^{n-1} remains in S_0^{n-1}, this yields

$$\begin{aligned}\langle Ax_0, y_0\rangle &= \langle A(x_0 - x_0'), y_0\rangle + \langle Ax_0', y_0 - y_0'\rangle + \langle Ax_0', y_0'\rangle \\ &\le a\|x_0 - x_0'\| + a\|y_0 - y_0'\| + \sup_{(x,y)\in\mathcal{N}_\varepsilon\times\mathcal{N}_\varepsilon^0} |\langle Ax, y\rangle|.\end{aligned}$$

Hence,

$$a \le 2\varepsilon\, a + R,$$

which gives that $a \le R/(1-2\varepsilon)$.

□

Now, we can prove the main statement of this section. It is easy to check that the theorem below, together with Proposition 4.5, gives Theorem C from the Introduction. To make the statement self-contained, we explicitly mention all the assumptions on parameters.

Theorem 5.5. *For any $L, m \ge 1$ there exist $\kappa = \kappa(L, m) > 0$ and $n_0 = n_0(L, m)$ with the following properties. Assume that $n \ge n_0$ and that the degree sequences $\mathbf{d}^{in}, \mathbf{d}^{out}$ satisfy*

$$(1 - c_0)d \le \mathbf{d}_i^{in}, \mathbf{d}_i^{out} \le d, \quad i \le n$$

for some natural d with $C_{3.2} \ln^2 n \le d \le (1/2 + c_0)n$. Then, with $\mathcal{E}_{\mathcal{P}}(L)$ defined by (4), we have

$$\mathbb{P}\big\{M \in \mathcal{M}_n(\mathbf{d}^{in}, \mathbf{d}^{out}) : s_2(M) \ge \kappa\sqrt{d}\big\} \le \frac{1}{n^m} + \mathbb{P}(\mathcal{E}_{\mathcal{P}}(L)^c).$$

Proof. Let $K_1 = K_1(L, m+1)$ and $K_2 = K_2(L, m+1)$ be defined as in Proposition 5.2, and let $\gamma = \gamma(L)$ and $\beta = \beta(K_1, K_2)$ be functions from Lemmas 5.1

and 5.3. We will use the shorter notation $\mathcal{E}_{\mathcal{P}}$ and $\mathcal{E}_{5.2}$ instead of $\mathcal{E}_{\mathcal{P}}(L)$ and $\mathcal{E}_{5.2}(K_1, K_2)$, respectively. Set

$$r := \gamma^{-1} H^{-1}(1 + \ln 81)$$

and denote

$$\mathcal{E} := \big\{M \in \mathcal{M}_n(\mathbf{d}^{in}, \mathbf{d}^{out}) : \ s_2(M) \geq 2(C_{5.1}\, L + \beta + r)\sqrt{d}\big\}.$$

Using the Courant–Fischer formula, we obtain

$$\mathbb{P}(\mathcal{E} \mid \mathcal{E}_{\mathcal{P}} \cap \mathcal{E}_{5.2}) \leq \mathbb{P}\Big\{ \exists (x,y) \in S^{n-1} \times S_0^{n-1} \text{ such that } |\langle \mathbf{M}y, x\rangle| \geq 2(C_{5.1}\, L + \beta + r)\sqrt{d} \mid \mathbf{M} \in \mathcal{E}_{\mathcal{P}} \cap \mathcal{E}_{5.2}\Big\}.$$

Let $\mathcal{N}$ be a 1/4-net in S^{n-1} and $\mathcal{N}_0$ be a 1/4-net in S_0^{n-1}. Standard volumetric estimates show that we may take $\mathcal{N}$ and $\mathcal{N}_0$ such that $\max(|\mathcal{N}|, |\mathcal{N}_0|) \leq 9^n$. Applying Lemma 5.4, we get

$$\begin{aligned}\mathbb{P}\{\mathcal{E} \mid \mathcal{E}_{\mathcal{P}} \cap \mathcal{E}_{5.2}\} &\leq \mathbb{P}\Big\{ \exists (x,y) \in \mathcal{N} \times \mathcal{N}_0 \text{ such that } |\langle \mathbf{M}y, x\rangle| \geq (C_{5.1}\, L + \beta + r)\sqrt{d} \mid \mathbf{M} \in \mathcal{E}_{\mathcal{P}} \cap \mathcal{E}_{5.2}\Big\} \\ &\leq (81)^n \max_{(x,y)\in S^{n-1}\times S_0^{n-1}} \mathbb{P}\Big\{ |\langle \mathbf{M}y, x\rangle| \geq (C_{5.1}\, L + \beta + r)\sqrt{d} \mid \mathbf{M} \in \mathcal{E}_{\mathcal{P}} \cap \mathcal{E}_{5.2}\Big\}.\end{aligned} \tag{39}$$

Given $(x,y) \in S^{n-1} \times S_0^{n-1}$, we obviously have

$$|\langle \mathbf{M}y, x\rangle| \leq \Big| \sum_{(i,j)\in\mathcal{L}(x,y)} x_i \mathbf{M}_{ij} y_j \Big| + \Big| \sum_{(i,j)\in\mathcal{H}(x,y)} x_i \mathbf{M}_{ij} y_j \Big|.$$

From Lemma 5.3, we get $\Big|\sum_{(i,j)\in\mathcal{H}(x,y)} x_i \mathbf{M}_{ij} y_j\Big| \leq \beta\sqrt{d}$ whenever $\mathbf{M} \in \mathcal{E}_{5.2}$. Hence, in view of (39),

$$\mathbb{P}(\mathcal{E} \mid \mathcal{E}_{\mathcal{P}}\cap\mathcal{E}_{5.2}) \leq (81)^n \max_{(x,y)\in S^{n-1}\times S_0^{n-1}} \mathbb{P}\Big\{\Big| \sum_{(i,j)\in\mathcal{L}(x,y)} x_i \mathbf{M}_{ij} y_j \Big| \geq (C_{5.1}\, L{+}r)\sqrt{d} \mid \mathcal{E}_{\mathcal{P}}\cap\mathcal{E}_{5.2}\Big\}.$$

Applying Lemma 5.1, we further obtain, by the choice of r,

$$\mathbb{P}(\mathcal{E} \mid \mathcal{E}_{\mathcal{P}} \cap \mathcal{E}_{5.2}) \leq \frac{C_{5.1}\,(81)^n}{\mathbb{P}(\mathcal{E}_{\mathcal{P}})} \exp\left(-n\, H(\gamma\, r)\right) \leq \frac{C_{5.1}\, e^{-n}}{\mathbb{P}(\mathcal{E}_{\mathcal{P}})}.$$

To finish the proof, note that

$$\mathbb{P}(\mathcal{E}) \leq \mathbb{P}(\mathcal{E} \mid \mathcal{E}_{\mathcal{P}} \cap \mathcal{E}_{5.2})\, \mathbb{P}(\mathcal{E}_{\mathcal{P}}) + \mathbb{P}(\mathcal{E}_{5.2}^c \mid \mathcal{E}_{\mathcal{P}})\, \mathbb{P}(\mathcal{E}_{\mathcal{P}}) + \mathbb{P}(\mathcal{E}_{\mathcal{P}}^c)$$

and use the above estimate together with Proposition 5.2. □

The concentration inequality obtained in Theorem 4.8, was used in its full strength in Proposition 5.2 to control the input of heavy couples. For the light couples though, it would be sufficient to apply a weaker Bernstein–type bound where the function $H(\tau)$ in the exponent is replaced with $\frac{\tau^2}{2+2\tau/3}$.

6. The undirected setting

In this section, we show how to deduce Theorem A from Theorem C. In [26], we showed that in a rather general setting the norm of a random matrix, whose distribution is invariant under joint permutations of rows and columns, can be bounded in terms of the norm of its $n/2 \times n/2$ submatrix located in the top right corner. Moreover, for matrices with constant row and column sums, an analogous phenomenon holds for the second largest singular values. Since the distribution of edges in the undirected uniform model is invariant under permutation of the set of vertices, the results of [26] are applicable in our context.

We will need the following definition. For any $\ell, d > 0$ and any parameter $\delta > 0$ we set

$$\begin{aligned}\mathbf{Deg}_\ell(d,\delta) := \Big\{(u,v) \in \mathbb{N}^\ell \times \mathbb{N}^\ell :\ &\|u\|_1 = \|v\|_1 \quad \text{AND} \\ &\big|\big\{i \leq \ell :\ \big|u_i - d\big| > k\delta\big\}\big| \leq \ell e^{-k^2} \text{ for all } k \in \mathbb{N} \quad \text{AND} \\ &\big|\big\{i \leq \ell :\ \big|v_i - d\big| > k\delta\big\}\big| \leq \ell e^{-k^2} \text{ for all } k \in \mathbb{N}\Big\}.\end{aligned}$$

Note that any pair of vectors (u, v) from $\mathbf{Deg}_\ell(d, \delta)$ necessarily satisfy $\|u - d\mathbf{1}\|_{\psi,n}, \|v - d\mathbf{1}\|_{\psi,n} \leq C\delta$ for some universal constant $C > 0$.

Below we state a special case of the main result of [26], where we replace a general random matrix with constant row/column sums by the adjacency matrix of a random regular graph.

Theorem 6.1 ([26])**.** *There exist positive universal constants c, C such that the following holds. Let $n \geq C$ and let $d \in \mathbb{N}$ satisfy $d \geq C \ln n$. Further, let $\mathbf{G}$ be a random undirected graph uniformly distributed on $\mathcal{G}_n(d)$ and let T be the $\lfloor n/2 \rfloor \times \lfloor n/2 \rfloor$ top right corner of the adjacency matrix of $\mathbf{G}$. Then, viewing T as the adjacency matrix of a random directed graph on $\lfloor n/2 \rfloor$ vertices, for any $t \geq C$ we have*

$$\mathbb{P}\big\{s_2(\mathbf{G}) \geq Ct\sqrt{d}\big\} \leq \frac{1}{c}\mathbb{P}\Big\{s_2(T) \geq ct\sqrt{d} \ \ AND \ \big(\mathbf{d}^{in}(T), \mathbf{d}^{out}(T)\big) \in \mathbf{Deg}_{\lfloor n/2 \rfloor}\big(d/2, C\sqrt{d}\big)\Big\}.$$

Equipped with the above statement and with Theorem C, we can proceed with the proof of Theorem A.

Proof of Theorem A. Let $m \in \mathbb{N}$, $\alpha > 0$ and let c, C be the constants from Theorem 6.1. We assume that $n^\alpha \leq d \leq n/2$. Denote by $A = (a_{ij})$ the adjacency matrix of the random graph $\mathbf{G}$ uniformly distributed on $\mathcal{G}_n(d)$. Let T be the $\lfloor n/2 \rfloor \times \lfloor n/2 \rfloor$ top right corner of A.

Fix for a moment *any* degree sequences $(\mathbf{d}^{in}, \mathbf{d}^{out})$ of length $\lfloor n/2 \rfloor$ bounded above by d such that the event $\{(\mathbf{d}^{in}(T), \mathbf{d}^{out}(T)) = (\mathbf{d}^{in}, \mathbf{d}^{out})\}$ is non-empty. Then, conditioned on the event, the directed random graph on $\lfloor n/2 \rfloor$ vertices with adjacency matrix T is *uniformly* distributed on $\mathcal{D}_{\lfloor n/2 \rfloor}(\mathbf{d}^{in}, \mathbf{d}^{out})$. In other words, the distribution of T, conditioned on the event $\{(\mathbf{d}^{in}(T), \mathbf{d}^{out}(T)) = (\mathbf{d}^{in}, \mathbf{d}^{out})\}$, is uniform on the set $\mathcal{M}_{\lfloor n/2 \rfloor}(\mathbf{d}^{in}, \mathbf{d}^{out})$.

Now if $(\mathbf{d}^{in}, \mathbf{d}^{out}) \in \mathbf{Deg}_{\lfloor n/2 \rfloor}(d/2, C\sqrt{d})$, then, applying Theorem C, we get

$$\mathbb{P}\Big\{s_2(T) \geq \tilde{t}\sqrt{d} \mid \big(\mathbf{d}^{in}(T), \mathbf{d}^{out}(T)\big) = (\mathbf{d}^{in}, \mathbf{d}^{out})\Big\} \leq \frac{1}{n^m}, \tag{40}$$

for some $\tilde{t}$ depending on α, C and m. Set $t := C\max(1, \tilde{t}/c)$. In view of Theorem 6.1, we get

$$\mathbb{P}\big\{s_2(\mathbf{G}) \geq t\sqrt{d}\big\} \leq \frac{1}{c}\mathbb{P}\Big\{s_2(T) \geq \tilde{t}\sqrt{d} \ \text{ AND } \big(\mathbf{d}^{in}(T), \mathbf{d}^{out}(T)\big) \in \mathbf{Deg}_{\lfloor n/2 \rfloor}\big(d/2, C\sqrt{d}\big)\Big\} =: \eta.$$

Obviously,

$$\eta = \frac{1}{c} \sum_{(\mathbf{d}^{in}, \mathbf{d}^{out}) \in \mathbf{Deg}_{\lfloor n/2 \rfloor}\left(d/2, C\sqrt{d}\right)} \mathbb{P}\Big\{s_2(T) \geq \tilde{t}\sqrt{d} \ \text{ AND } \big(\mathbf{d}^{in}(T), \mathbf{d}^{out}(T)\big) = (\mathbf{d}^{in}, \mathbf{d}^{out})\Big\}.$$

Hence, applying (40), we get

$$\eta \leq \frac{1}{c\,n^m} \sum_{(\mathbf{d}^{in}, \mathbf{d}^{out}) \in \mathbf{Deg}_{\lfloor n/2 \rfloor}\left(d/2, C\sqrt{d}\right)} \mathbb{P}\Big\{\big(\mathbf{d}^{in}(T), \mathbf{d}^{out}(T)\big) = (\mathbf{d}^{in}, \mathbf{d}^{out})\Big\} \leq \frac{1}{c\,n^m},$$

and complete the proof. □

Acknowledgments. A significant part of this work was done when the second named author visited the University of Alberta in May–June 2016, and when both authors visited the Texas A&M University in July 2016. Both authors are grateful to the University of Alberta and the Texas A&M University for excellent working conditions, and would especially like to thank Nicole Tomczak–Jaegermann, Bill Johnson, Alexander Litvak and Grigoris Paouris. We would also like to thank Djalil Chafaï for helpful comments. The first named author is partially supported by the Simons Foundation (*Collaboration on Algorithms and Geometry*).

References

[1] N. Alon, Eigenvalues and expanders, Combinatorica **6** (1986), no. 2, 83–96. MR0875835

[2] N. Alon and V. D. Milman, λ_1, isoperimetric inequalities for graphs, and superconcentrators, J. Combin. Theory Ser. B **38** (1985), no. 1, 73–88. MR0782626

[3] R. Bauerschmidt, J. Huang, A.Knowles, and H.-T. Yau. Bulk eigenvalue statistics for random regular graphs, arXiv:1505.06700.

[4] R. Bauerschmidt, A. Knowles, H.-T. Yau, Local semicircle law for random regular graphs, arXiv:1503.08702.
[5] G. Bennett, Probability Inequalities for the Sum of Independent Random Variables, Journal of the American Statistical Association **297** (1962), 33–45, doi:10.2307/2282438.
[6] S. N. Bernstein, Theory of Probability (in Russian), Moscow, 1927.
[7] C. Bordenave, A new proof of Friedman's second eigenvalue Theorem and its extension to random lifts, arXiv:1502.04482.
[8] A. Z. Broder, A. M. Frieze, S. Suen, E. Upfal, Optimal construction of edge-disjoint paths in random graphs, SIAM J. Comput. **28** (1999), no. 2, 541–573 (electronic). MR1634360
[9] A. Broder, E. Shamir, On the second eigenvalue of random regular graphs, Proceedings of the 28th Annual Symposium on Foundations of Computer Science (1987), 286–294.
[10] N. Cook, Discrepancy properties for random regular digraphs, Random Structures Algorithms, DOI: 10.1002/rsa.20643.
[11] N. Cook, On the singularity of adjacency matrices for random regular digraphs, Prob. Theory and Related Fields, to appear. arXiv:1411.0243.
[12] N. Cook, L. Goldstein, T. Johnson, Size biased couplings and the spectral gap for random regular graphs, arXiv:1510.06013.
[13] I. Dumitriu, T. Johnson, S. Pal, E. Paquette, Functional limit theorems for random regular graphs, Probab. Theory Related Fields **156** (2013), no. 3-4, 921–975. MR3078290
[14] I. Dumitriu and S. Pal, Sparse regular random graphs: spectral density and eigenvectors, Ann. Probab. **40** (2012), no. 5, 2197–2235. MR3025715
[15] D. A. Freedman, On tail probabilities for martingales, Ann. Probability **3** (1975), 100–118. MR0380971
[16] J. Friedman, A proof of Alon's second eigenvalue conjecture and related problems, Mem. Amer. Math. Soc. **195** (2008), no. 910, viii+100 pp. MR2437174
[17] J. Friedman, On the second eigenvalue and random walks in random d-regular graphs, Combinatorica **11** (1991), no. 4, 331–362. MR1137767
[18] J. Friedman, J. Kahn, E. Szemerédi, On the second eigenvalue of random regular graphs, Proceedings of the twenty-first annual ACM symposium on Theory of computing (1989), 587–598.
[19] A.E. Litvak, A. Lytova, K. Tikhomirov, N. Tomczak-Jaegermann, P. Youssef, Adjacency matrices of random digraphs: singularity and anti-concentration, J. of Math. Analysis and Appl., 445 (2017), 1447-1491. arXiv:1511.00113.
[20] A. E. Litvak, A. Pajor, M. Rudelson, N. Tomczak-Jaegermann, Smallest singular value of random matrices and geometry of random polytopes. Adv. Math. 195 (2005), no. 2, 491-523.
[21] B. D. McKay, The expected eigenvalue distribution of a large regular graph, Linear Algebra Appl. **40** (1981), 203–216. MR0629617
[22] D. Puder. Expansion of random graphs: New proofs, new results, Inventiones Mathematicae, 201 (3), 845-908, 2015.

[23] M. M. Rao and Z. D. Ren, *Theory of Orlicz spaces*, Monographs and Textbooks in Pure and Applied Mathematics, 146, Dekker, New York, 1991. MR1113700

[24] M. Rudelson and R. Vershynin, The Littlewood-Offord problem and invertibility of random matrices. Adv. Math. 218 (2008), no. 2, 600-633.

[25] J. K. Senior, Partitions and their representative graphs, Amer. J. Math. **73** (1951), 663–689. MR0042678

[26] K. Tikhomirov and P. Youssef, On the norm of a random jointly exchangeable matrix, arXiv:1610.01751.

[27] L. V. Tran, V. H. Vu and K. Wang, Sparse random graphs: eigenvalues and eigenvectors, Random Structures Algorithms **42** (2013), no. 1, 110–134. MR2999215

[28] V. Vu, Random discrete matrices, in *Horizons of combinatorics*, 257–280, Bolyai Soc. Math. Stud., 17, Springer, Berlin. MR2432537

[29] V. Vu, Combinatorial problems in random matrix theory, Proceedings ICM, Vol. 4, 2014, 489–508.

[30] N. C. Wormald, Models of random regular graphs, in *Surveys in combinatorics, 1999 (Canterbury)*, 239–298, London Math. Soc. Lecture Note Ser., 267, Cambridge Univ. Press, Cambridge. MR1725006