Received: 20 January 2020 | Revised: 23 October 2021 | Accepted: 13 March 2022 | Published on: 23 April 2022

DOI: 10.1002/rsa.21085

RESEARCH ARTICLE

WILEY

The rank of sparse random matrices

Amin Coja-Oghlan[1] | **Alperen A. Ergür**[2] | **Pu Gao**[3] | **Samuel Hetterich**[4] | **Maurice Rolvien**[1]

[1]Faculty of Computer Science, TU Dortmund, Dortmund, Germany
[2]The University of Texas at San Antonio, San Antonio, Texas, USA
[3]Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada
[4]Mathematics Institute, Goethe University, Frankfurt, Germany

Correspondence
Amin Coja-Oghlan, Faculty of Computer Science, TU Dortmund, 12 Otto Hahn St, Dortmund 44227, Germany.
Email: amin.coja-oghlan@tu-dortmund.de

Abstract

We determine the asymptotic normalized rank of a random matrix $\boldsymbol{A}$ over an arbitrary field with prescribed numbers of nonzero entries in each row and column. As an application we obtain a formula for the rate of low-density parity check codes. This formula vindicates a conjecture of Lelarge (2013). The proofs are based on coupling arguments and a novel random perturbation, applicable to any matrix, that diminishes the number of short linear relations.

KEYWORDS

finite field, random constraint satisfaction, random matrices, rank, sparse matrices

1 | INTRODUCTION

1.1 | Background and motivation

The theory of random matrices, which commenced with the nuclear physics-inspired work of Wigner in the 1950s [56], has been one of the great success stories at the junction of probability, mathematical physics and combinatorics. Nevertheless, quite a few basic questions remain open to this day. For instance, while dense random matrices such as the Gaussian Orthogonal Ensemble are reasonably well understood (e.g., [30]), far less is known about sparse random matrices where the expected number of nonzero entries per row or column is bounded. Yet over the last two or three decades such sparse random matrices, with entries from finite or infinite fields, have emerged to play a pivotal role in several exciting applications. Modern error-correcting codes are a case in point. For instance, the codebook of a low-density parity check code ("ldpc code"), a class of codes that has been at the centre of tremendous recent developments in coding theory [20, 27, 34], comprises the kernel of a sparse random matrix over

a finite field drawn from a carefully tailored distribution. In addition, sparse random matrices occur in randomised constructions of Ramanujan graphs [6, 13, 19], statistical inference [26], the analysis of algorithms [16], and the theory of random constraint satisfaction problems [1, 23].

Among the fundamental questions about such random matrices that have remained open, perhaps the most conspicuous one concerns the rank. Although this parameter was already studied in early contributions [8, 9, 37], there has been no comprehensive rank formula for sparse random matrices. The present paper furnishes one. To be precise, we will determine the asymptotic rank of a sparse random matrix with prescribed numbers of nonzero entries in the rows and columns. Among other applications, important classes of ldpc codes are based on precisely such random matrices as a diligent choice of the degrees greatly boosts the code's performance [34]. Moreover, the rank is linearly related to the rate of the code, arguably the code's most basic parameter.

Lelarge [41] noticed that an upper bound on the rank of a sparse random matrix can be derived from the matching number of random bipartite graphs, which was determined by Bordenave, Lelarge, and Salez [14]. Lelarge went on to conjecture that this bound be tight for sparse random matrices over the binary field $\mathbb{F}_2$. We prove this conjecture. In fact, we prove a much stronger result. Namely, we show that Lelarge's conjectured formula holds for sparse random matrices over any field, finite or infinite, regardless the distribution of the nonzero matrix entries. Thus, the rank is governed by the *location* of the nonzero entries rather than the distribution of the matrix entries.

The proof of the rank formula evinces an interesting connection to statistical physics. Indeed, Lelarge already observed that a sophisticated but mathematically nonrigorous physics approach called the "cavity method" renders a wrong prediction as to the rank for certain degree distributions.[1] This discrepancy merits attention because the cavity method has been brought to bear on a panoply of theoretical as well as real-world problems, ranging from spin glasses to machine learning [57]. We manage to shed light on the issue. Specifically, the "replica symmetric" version of the cavity method predicts that the rank of a random matrix over a finite field can be expressed analytically as the maximum of a variational problem. A priori, this variational problem asks to optimize a functional called the Bethe free entropy over an infinite-dimensional space of probability measures. Such optimization problems have been tackled in the physics literature numerically by means of a heuristic called population dynamics. For the rank problem this was carried out by Alamino and Saad [3]. But thanks to the algebraic nature of the problem we can show that the rank actually comes out as the solution to a variational problem on a restricted domain. We are thus left with a dramatically simplified variational problem, which ultimately boils down to a humble one-dimensional optimization task. We will see that the optimal solution to this one-dimensional problem does indeed yield the rank (over any field). Furthermore, the solution can be lifted to a solution to the original infinite-dimensional problem. As an aside, we do not know if the original infinite-dimensional variational problem may possess spurious maximizers that boost its value beyond the optimal value of the restricted version, thereby spoiling the accuracy of the original physics formula. We will return to this question, and to the physics slant on the problem, in Section 2.3.3. In any case, for certain degree distributions the maximum values that we obtain by way of the restricted variational problem actually exceed those that surfaced in the experiments from [3] or the heuristic derivations from [31] for the unrestricted formula; hence the discrepancy between the physics predictions and mathematical reality.

Apart from remedying the discrepancy, we prove the rank formula by effectively turning the physicists' cavity calculations into a rigorous mathematical argument. The crucial tool that makes this possible is a novel perturbation, applicable to any matrix, that diminishes the number of short linear relations (see Proposition 2.4 below). We expect that this perturbation will find future applications. Let

[1]The derivation of this erroneous prediction was posed as an exercise in [31, Chapter 19].

us proceed to introduce the random matrix model and state the main results. A discussion of related work and a detailed comparison with the physics work follow in Section 2, once we have the necessary notation in place.

1.2 | The rank formula

Let $\mathbb{F}$ be a field equipped with a σ-algebra that turns $\mathbb{F}$ into a standard Borel space and let $\chi : [0,1]^2 \to \mathbb{F}^* = \mathbb{F} \setminus \{0\}$ be a measurable map. Let $(\boldsymbol{\zeta}_i, \boldsymbol{\xi}_i)_{i\geq 1}$ be mutually independent uniformly distributed $[0,1]$-valued random variables. Moreover, let $\boldsymbol{d}, \boldsymbol{k} \geq 0$ be integer-valued random variables such that $0 < \mathbb{E}[\boldsymbol{d}^r] + \mathbb{E}[\boldsymbol{k}^r] < \infty$ for a real $r > 2$ and set $d = \mathbb{E}[\boldsymbol{d}]$, $k = \mathbb{E}[\boldsymbol{k}]$. Let $n > 0$ be an integer divisible by the greatest common divisor of the support of $\boldsymbol{k}$ and let $\boldsymbol{m} \sim \mathrm{Po}(dn/k)$ be independent of the $\boldsymbol{\zeta}_i, \boldsymbol{\xi}_i$. Further, let $(\boldsymbol{d}_i, \boldsymbol{k}_i)_{i\geq 1}$ be copies of $\boldsymbol{d}, \boldsymbol{k}$, mutually independent and independent of $\boldsymbol{m}, \boldsymbol{\zeta}_i, \boldsymbol{\xi}_i$. Given

$$\sum_{i=1}^{n} \boldsymbol{d}_i = \sum_{i=1}^{\boldsymbol{m}} \boldsymbol{k}_i, \tag{1.1}$$

draw a simple bipartite graph $\boldsymbol{G}$ comprising a set $\{a_1, \ldots, a_{\boldsymbol{m}}\}$ of *check nodes* and a set $\{x_1, \ldots, x_n\}$ of *variable nodes* such that the degree of a_i equals $\boldsymbol{k}_i$ and the degree of x_j equals $\boldsymbol{d}_j$ for all i, j uniformly at random. Then let $\boldsymbol{A}$ be the $\boldsymbol{m} \times n$-matrix with entries

$$\boldsymbol{A}_{ij} = \mathbf{1}\{a_i x_j \in E(\boldsymbol{G})\} \cdot \chi_{\zeta_i, \xi_j}.$$

Thus, the ith row of $\boldsymbol{A}$ features precisely $\boldsymbol{k}_i$ nonzero entries and the jth column contains precisely $\boldsymbol{d}_j$ nonzero entries. Moreover, the nonzero entries of $\boldsymbol{A}$ are drawn in the vein of an exchangeable array by evaluating the function χ at a random poisition $(\boldsymbol{\zeta}_i, \boldsymbol{\xi}_j)$. Routine arguments show that $\boldsymbol{A}$ is well-defined for large enough n, that is, (1.1) is satisfied and there exists a simple $\boldsymbol{G}$ with the desired degrees with positive probability; see Proposition 1.10 below. We call $\boldsymbol{G}$ the *Tanner graph*of $\boldsymbol{A}$. Also recall that the rank $\mathrm{rk}\, A$ of the matrix A is defined as the maximal number of linear independent rows (or columns). In addition, $\mathrm{nul}\, A$ is the dimension of the kernel of A and the sum $\mathrm{rk}\, A + \mathrm{nul}\, A$ equals the number of columns of A.

The following theorem, the main result of the paper, provides an asymptotic formula for the rank of $\boldsymbol{A}$. Let $D(x)$ and $K(x)$ denote the probability generating functions of $\boldsymbol{d}$ and $\boldsymbol{k}$, respectively. Since $\mathbb{E}[\boldsymbol{d}^2] + \mathbb{E}[\boldsymbol{k}^2] < \infty$, the functions $D(x), K(x)$ are continuously differentiable on the unit interval. Therefore, the function

$$\Phi : [0,1] \to \mathbb{R}, \alpha \mapsto D\left(1 - K'(\alpha)/k\right) - \frac{d}{k}\left(1 - K(\alpha) - (1-\alpha)K'(\alpha)\right). \tag{1.2}$$

is continuous.

Theorem 1.1. *For any $\boldsymbol{d}, \boldsymbol{k}$ we have, uniformly for all χ,*

$$\lim_{n\to\infty} \frac{\mathrm{rk}(\boldsymbol{A})}{n} = 1 - \max_{\alpha\in[0,1]} \Phi(\alpha) \quad \text{in } \textit{probability}. \tag{1.3}$$

Perhaps surprisingly, the r.h.s. of (1.3) depends only on the degree distributions $\boldsymbol{d}, \boldsymbol{k}$ but not in any way on the field $\mathbb{F}$ or the choice of nonzero entries (within the aforementioned model). Furthermore, let us emphasize that the function Φ, being continuous on the unit interval, is guaranteed to attain a maximum. However, this maximum need not be unique, and nonuniqueness of the maximizer may have interesting combinatorial repercussions [16].

A second point that may seem surprising at first glance is that the rank converges to any nonrandom value at all, as provided by (1.3). A heuristic explanation can be given on grounds of physics reasoning. Indeed, the nullity of $\boldsymbol{A}$ (dimension of the kernel) corresponds to the logarithm of the partition function of a natural Boltzmann distribution, namely the uniform distribution on the kernel of $\boldsymbol{A}$. Commonly the normalized logarithm of such a partition functions (known as the "free entropy" in physics jargon) converges to a constant for random systems that are "self-averaging." Here "self-averaging" means that a small perturbation to the system, that is, the matrix in our case, cannot cause disproportionate tremors in logarithm of the partition function. In the random matrix model that we consider here the self-averaging condition is clearly satisfied because changing a single matrix entry can at most alter the nullity by one. Therefore, the Azuma–Hoeffding inequality easily implies that $\operatorname{nul}\boldsymbol{A}$ concentrates about its mean. That said, there is no general theorem that guarantees convergence to a deterministic value in self-averaging systems, so even this aspect of Theorem 1.1 is not in any way a triviality.

Theorem 1.1 establishes a generalised version of Lelarge's rank conjecture [41] with a tighter conditions on the moments of $\boldsymbol{d},\boldsymbol{k}$. Specifically, Lelarge only considered matrices over the field $\mathbb{F}_2$, while here we consider general fields and allow for a very general choice of nonzero entries. That said, while here we assume that $\mathbb{E}[\boldsymbol{d}^r],\mathbb{E}[\boldsymbol{k}^r]<\infty$ for a real $r>2$, Lelarge considered degree distributions with $\mathbb{E}[\boldsymbol{d}^2],\mathbb{E}[\boldsymbol{k}^2]<\infty$. We did not undertake a serious attempt to weaken the moment condition to $r=2$, but this may conceivably introduce significant new techical difficulties.

The theorem covers a very general class of sparse random matrices. Indeed, since $\boldsymbol{d},\boldsymbol{k}$ have finite means the matrix $\boldsymbol{A}$ is sparse, that is, the expected number of nonzero entries is $O(n)$ as $n\to\infty$. Yet because the degree distributions are subject only to the condition $\mathbb{E}[\boldsymbol{d}^r]+\mathbb{E}[\boldsymbol{k}^r]<\infty$, the typical maximum number of nonzero entries per row or column may approach $\sqrt{n}$. Furthermore, the choice of the nonzero entries of the matrix by way of the measurable map χ, reminiscent of an exchangeable array, allows for rather general choices of nonzero matrix entries. To elaborate, recall that an exchangeable array is an infinite matrix $(\chi_{ij})_{i,j\geq1}$ of $\mathbb{F}^*$-valued random variables such that the distribution of any finite top-left submatrix is invariant under row and column permutations [35]. The Aldous–Hoover representation theorem shows that any such array can be described by a function $\mathcal{X}:[0,1]^4\to\mathbb{F}^*$ [4, 31]. Specifically, any finite submatrix of χ_{ij} can be obtained by substituting suitable independent random variables that are uniformly distributed on the unit interval $[0,1]$ into $\mathcal{X}$. Theorem 1.1 therefore implies the rank formula for a Hadamard product of the biadjacency matrix of the random bipartite graph $\boldsymbol{G}$ and the commensurately dimensioned top-left bit of the exchangeable array $(\chi_{ij})_{i,j}$. Of course, an immediate special case is the random matrix whose nonzero entries are drawn mutually independently from an arbitrary distribution on $\mathbb{F}^*$. [2]

The lower bound on the rank constitutes the principal contribution of Theorem 1.1. Indeed, the upper bound $\operatorname{rk}(\boldsymbol{A})/n\leq 1-\max_{\alpha\in[0,1]}\Phi(\alpha)+o(1)$ as $n\to\infty$ a.a.s. was already derived in [41] from the Leibniz determinant formula and the formula for the matching number of a random bipartite graph from [14].[3] Nonetheless, in the appendix we give an independent proof of the upper bound, which is shorter than the combination [14, 41].

Theorem 1.1 implies a formula for the rate of a common class of ldpc codes. Such codes are based on random matrices $\boldsymbol{A}$ over finite fields $\mathbb{F}_q$ with suitable degree distributions $\boldsymbol{d},\boldsymbol{k}$. Specifically, a common construction of ldpc codes involves an optimisation over the degree distributions $\boldsymbol{d},\boldsymbol{k}$ of

[2]To see this, assume that χ is an $\mathbb{F}^*$-valued random variable. Then given n pick a large integer $N\gg n^2$. Let $\chi:[0,1]^2\to\mathbb{F}^*$ be a step function obtained by chopping $[0,1]$ into N subintervals of size $1/N$ and assigning a value drawn from χ independently to each of the N^2 resulting rectangles. Because Theorem 1.1 provides uniform convergence in χ, we obtain the rank of a matrix with nonzero entries drawn from χ.

[3]While [41] only dealt with matrices over $\mathbb{F}_2$, the argument extends to other fields without further ado.

the variables/checks so as to maximise the probability that the Belief Propagation message passing algorithm (or a variant thereof) recovers the original codeword from the received, noisy data [52]. The codebook consists of the kernel of the random matrix $\boldsymbol{A}$. Hence, the rate of the code equals $\operatorname{nul}\boldsymbol{A}/n$. Since Theorem 1.1 implies that

$$\frac{1}{n}\operatorname{nul}\boldsymbol{A} \to \max_{\alpha\in[0,1]}\Phi(\alpha) \quad \text{in probability},$$

we thus obtain the rate.

1.3 | The 2-core bound

There is a simple graph-theoretic upper bound on the rank, and Theorem 1.1 puts us in a position to investigate if and when this bound is tight. To state this bound, we recall that the *2-core* of $\boldsymbol{G}$ is the subgraph $\boldsymbol{G}_*$ obtained by repeating the following operation.

> While there is a variable node x_i of degree one or less, remove that variable node along with the adjacent check node (if any).[4]

Of course, the 2-core may be empty, that is, with no variable or check nodes. In the case that $\mathbb{P}(\boldsymbol{k}=0)>0$ it is possible to have a 2-core without any variable node but with a non-empty set of check nodes whose degrees are all zero. Extending prior results that dealt with the degrees of all check nodes coinciding [19, 47], we compute the likely number of variable and check nodes in the 2-core. Let

$$\phi(\alpha) = 1-\alpha-D'\left(1-K'(\alpha)/k\right)/d. \tag{1.4}$$

Note that $\Phi'(\alpha) = dK''(\alpha)\phi(\alpha)/k$. Since $\boldsymbol{d},\boldsymbol{k}$ have finite second moments and $\phi(0)\geq 0$ while $\phi(1)\leq 0$, we can define

$$\rho = \max\{x\in[0,1]:\phi(x)=0\}. \tag{1.5}$$

Theorem 1.2. *Assume that $\phi'(\rho)<0$ and let $\boldsymbol{n}^*$ and $\boldsymbol{m}^*$ be the number of variable and check nodes in the 2-core, respectively. Then*

$$\lim_{n\to\infty}\frac{\boldsymbol{n}^*}{n} = 1-D\left(1-\frac{K'(\rho)}{k}\right)-\frac{K'(\rho)}{k}D'\left(1-\frac{K'(\rho)}{k}\right), \quad \lim_{n\to\infty}\frac{\boldsymbol{m}^*}{n}=\frac{d}{k}K(\rho) \text{ in probability}. \tag{1.6}$$

Remark 1.3.

(a) If $\mathbb{P}(\boldsymbol{k}=1)=0$ then $1-D\left(1-\frac{K'(0)}{k}\right)-\frac{K'(0)}{k}D'\left(1-\frac{K'(0)}{k}\right)$ evaluates to zero, and a.a.s. $dK(0)/k$ is the number of check nodes with degree zero in $\boldsymbol{G}$ divided by n, up to an $o(1)$ error.

(b) If $\boldsymbol{d}\leq 1$ then we observe that $\phi(\alpha)=-\alpha$ and thus $\rho=0$. In this case, $1-D\left(1-\frac{K'(0)}{k}\right)-\frac{K'(0)}{k}D'\left(1-\frac{K'(0)}{k}\right)$ evaluates to zero. This agrees with the trivial fact that $\boldsymbol{n}^*=0$, and $\frac{\boldsymbol{m}^*}{n}\to dK(0)/k$ a.a.s. in this case.

[4] Strictly speaking, what we describe here is the 2-core of the hypergraph whose vertices are the variable nodes and whose edges are the neighborhoods of the check nodes.

(c) If $\mathbb{P}(\boldsymbol{k}=1)>0$ and $\mathbb{P}(\boldsymbol{d}\geq 2)>0$ then $\phi(0)>0$, which implies that $\rho>0$. Thus the right-hand sides of (1.6) are both positive.

Theorem 1.2 yields an elementary upper bound on the rank of $\boldsymbol{A}$, as follows, which we refer to as the *2-core bound*:

$$\mathrm{rk}(\boldsymbol{A})/n \leq 1-\max\{\Phi(0),\Phi(\rho)\}+o(1) \qquad \text{a.a.s.} \tag{1.7}$$

To see that $\mathrm{rk}(\boldsymbol{A})/n \leq 1-\Phi(0)+o(1)$ a.a.s., let $\boldsymbol{A}'$ be the matrix comprising the rows of $\boldsymbol{A}$ that contain at most one nonzero entry and let $\boldsymbol{m}'$ be the number of such rows. Then $\mathrm{rk}(\boldsymbol{A})\leq \boldsymbol{m}-\boldsymbol{m}'+\mathrm{rk}(\boldsymbol{A}')$. Moreover, routine arguments reveal that $(\boldsymbol{m}-\boldsymbol{m}')/n \sim d(1-K(0)-K'(0))/k$ and $\mathrm{rk}(\boldsymbol{A}')/n \sim 1-D(1-K'(0)/k)$ a.a.s. (see Appendix D for a proof), deducing the desired upper bound for $\mathrm{rk}(\boldsymbol{A})$.

The other upper bound in (1.7) can be deduced by considering the 2-core and lower bounding the nullity. Counting only solutions to $\boldsymbol{A}x=0$ where $x_i=0$ for all variables that belong to the 2-core $\boldsymbol{G}_*$, we obtain $\mathrm{nul}(\boldsymbol{A})\geq n-\boldsymbol{n}^*-(\boldsymbol{m}-\boldsymbol{m}^*)$. Invoking Theorem 1.2, we thus find that as $n\to\infty$,

$$\frac{\mathrm{rk}(\boldsymbol{A})}{n} \leq 1-D\left(1-\frac{K'(\rho)}{k}\right)+\frac{d}{k}(1-K(\rho))-\frac{K'(\rho)}{k}D'\left(1-\frac{K'(\rho)}{k}\right).$$

Now $\phi(\rho)=0$ implies $D'\left(1-\frac{K'(\rho)}{k}\right)=d(1-\rho)$. Substituting this into the inequality above yields

$$\mathrm{rk}(\boldsymbol{A})/n \leq 1-\Phi(\rho)+o(1) \quad \text{a.a.s.} \tag{1.8}$$

The following theorem shows that the 2-core bound is tight in several cases of interest.

Theorem 1.4. *Assume that*

(i) either $\mathrm{Var}(\boldsymbol{d})=0$ *or* $\boldsymbol{d}\sim \mathrm{Po}_{\geq \ell}(\lambda)$ *for an integer* $\ell\geq 0$ *and* $\lambda>0$, *and*
(ii) either $\mathrm{Var}(\boldsymbol{k})=0$ *or* $\boldsymbol{k}\sim \mathrm{Po}_{\geq \ell'}(\lambda')$ *for an integer* $\ell'\geq 0$ *and* $\lambda'>0$.

Then

$$\lim_{n\to\infty}\mathrm{rk}(\boldsymbol{A})/n = 1-\max\{\Phi(0),\Phi(\rho)\} \quad \text{in } \textit{probability}.$$

Remark 1.5. Under either condition of Theorem 1.4 (i) or (ii), the condition $\phi'(\rho)<0$ of Theorem 1.2 is satisfied, unless $\mathbb{P}(\boldsymbol{d}=1)=0$ and $2(k-1)\mathbb{P}(\boldsymbol{d}=2)\geq d$. We will prove this in the proof of Theorem 1.4.

On the basis of a canny but nonrigorous statistical physics approach called the cavity method several authors predicted that (over finite fields) the 2-core bound (1.7) is universally tight for all $\boldsymbol{d},\boldsymbol{k}$. Alamino and Saad reached this conclusion by way of numerical experiments [3], while Mézard and Montanari [46] posed a nonrigorous but analytical derivation as an exercise. However, the prediction turns out to be erroneous. Indeed, Lelarge [41] produced an example of $\boldsymbol{d},\boldsymbol{k}$ whose function $\Phi(\alpha)$ attains its unique maximum at a value $0<\alpha<\rho$. We will see another counter-example momentarily. On the positive side, Theorem 1.4 verifies that the 2-core bound actually is tight in all the cases for which Alamino and Saad [3] conducted numerical experiments.

1.4 | Examples

Let us conclude this section by investigating a few examples of degree distributions $\boldsymbol{d},\boldsymbol{k}$ and their resulting rank formulas.

Example 1.6 (the identity matrix)**.** As was brought to our attention by an anonymous reviewer, in the case $\boldsymbol{d} = \boldsymbol{k} = 1$ deterministically the matrix $\boldsymbol{A}$ is just a permutation matrix, which clearly has full rank. Accordingly, we find $D(x) = K(x) = x$ and $\Phi(x) = 0$. Hence, (1.3) boils down to the trivial fact $\mathrm{rk}\,\boldsymbol{A} \sim n$.

Example 1.7 (the adjacency matrix of random bipartite graphs)**.** Let $\mathbb{G} = \mathbb{G}(n,n,p)$ be a random bipartite graph on vertices $v_1, \ldots, v_n, v'_1, \ldots, v'_n$ such that for any $i,j \in [n]$ the edge $\{v_i, v'_j\}$ is present with probability p independently. With $p = \Delta/n$ for a fixed $\Delta > 0$ for large n the vertex degrees asymptotically have distribution $\mathrm{Po}(\Delta)$. Indeed, with the choice $\boldsymbol{d} \sim \mathrm{Po}(\Delta)$ and $\boldsymbol{k} \sim \mathrm{Po}(\Delta)$ the adjacency matrix $A(\mathbb{G}(n,n,p))$ and the random matrix $\boldsymbol{A}$ can be coupled such that $\mathrm{rk}\,A(\mathbb{G}(n,n,p)) = \mathrm{rk}(\boldsymbol{A}) + o(n)$ a.a.s. Hence, Theorem 1.1shows that over any field $\mathbb{F}$,

$$\lim_{n\to\infty} \frac{\mathrm{rk}(A(\mathbb{G}(n,n,p)))}{n} = 2 - \max\{\exp(-\Delta\exp(\Delta(\alpha-1))) + (1+(1-\alpha)\Delta)\exp(\Delta(\alpha-1)) : \alpha \in [0,1]\},$$

in probability. Theorem 1.4 implies that the 2-core bound is tight in this example.

Example 1.8 (fixed row sums)**.** Motivated by the minimum spanning tree problem on weighted random graphs, Cooper, Frieze and Pegden [20] studied the rank of the random matrix with degree distributions $\boldsymbol{k} = k \geq 3$ fixed and $\boldsymbol{d} \sim \mathrm{Po}(d)$ over the field $\mathbb{F}_2$. The same rank formula was obtained independently in [7] for arbitrary finite fields. Extending both these results, Theorem 1.1 shows that the rank of the random matrix with these degrees over any field $\mathbb{F}$ with any choice χ of nonzero entries is given by

$$\lim_{n\to\infty} \frac{\mathrm{rk}\,\boldsymbol{A}}{n} = 1 - \max\left\{\exp(-d\alpha^{k-1}) - \frac{d}{k}\left(1 - k\alpha^{k-1} + (k-1)\alpha^k\right) : \alpha \in [0,1]\right\}.$$

Once more Theorem 1.4 shows that the 2-core bound is tight.

Example 1.9 (nonexact 2-core bound)**.** There are plenty of choices of $\boldsymbol{d},\boldsymbol{k}$ where the 2-core bound fails to be tight. Degree distributions that render graphs $\boldsymbol{G}$ with an unstable 2-core furnish particularly egregious offenders. In such graphs the removal of a small number of randomly chosen checks a_i likely causes the 2-core to collapse. Analytically, the instability manifests itself in ρ from (1.5) being a local minimum of $\Phi(x)$. For instance, letting $\boldsymbol{d},\boldsymbol{k}$ be the distributions with $D(x) = (22x^2 + 3x^{11})/25$ and $K(x) = x^3$, we obtain

$$\Phi(x) = -\frac{3}{25}x^{22} + \frac{33}{25}x^{20} - \frac{33}{5}x^{18} + \frac{99}{5}x^{16} - \frac{198}{5}x^{14} + \frac{1386}{25}x^{12} - \frac{1386}{25}x^{10} + \frac{198}{5}x^8 - \frac{99}{5}x^6 + \frac{187}{25}x^4 - \frac{154}{75}x^3 - \frac{2}{75}. \tag{1.9}$$

Hence, $\rho = 1$ and $\Phi''(1) > 0$, while the global maximum is attained at $\alpha \approx 0.75$.

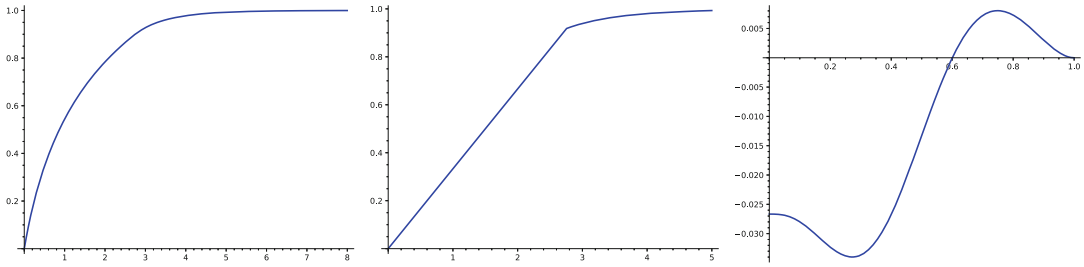

FIGURE 1 Left: the function $\Delta \mapsto 2 - \max_{\alpha\in[0,1]} \exp(-\Delta\exp(\Delta(\alpha-1))) + (1+(1-\alpha)\Delta)\exp(\Delta(\alpha-1))$ for Example 1.7. Middle: the function $d \mapsto 1 - \max_{\alpha\in[0,1]} \exp(-d\alpha^{k-1}) - d(1-k\alpha^{k-1}+(k-1)\alpha^k)/k$ from Example 1.8 with $k=3$. Right: the function $\Phi(x)$ from (1.9) for Example 1.9

1.5 | Preliminaries

Throughout the paper we consistently keep the assumptions on the distributions $\boldsymbol{d}, \boldsymbol{k}$ listed in Section 1. In particular, $\mathbb{E}[\boldsymbol{d}^r] + \mathbb{E}[\boldsymbol{k}^r] < \infty$ for some real $r > 2$. Because all-zero rows and columns do not add to the rank, we may assume that $\boldsymbol{d} \geq 1, \boldsymbol{k} \geq 1$. We write $\gcd(\boldsymbol{k})$ and $\gcd(\boldsymbol{d})$ for the greatest common divisor of the support of $\boldsymbol{d}$ and $\boldsymbol{k}$, respectively. When working with $\boldsymbol{A}$ we tacitly assume that $\gcd(\boldsymbol{k})$ divides n. In order to highlight the number of columns we write $\boldsymbol{A}_n = \boldsymbol{A}$ and $\boldsymbol{G}_n = \boldsymbol{G}$ for the corresponding Tanner graph. The following proposition, whose proof can be found in Section 4.2, shows that $\boldsymbol{A}_n$ is well-defined (Figure 1).

Proposition 1.10. *With probability $\Omega(n^{-1/2})$ over the choice of $\boldsymbol{m}$, $(\boldsymbol{d}_i)_{i\geq 1}$, $(\boldsymbol{k}_i)_{i\geq 1}$ the condition (1.1) is satisfied and there exists a simple Tanner graph $\boldsymbol{G}$ with variable degrees $\boldsymbol{d}_1, \ldots, \boldsymbol{d}_n$ and check degrees $\boldsymbol{k}_1, \ldots, \boldsymbol{k}_{\boldsymbol{m}}$.*

We introduce the size-biased random variables

$$\mathbb{P}\left[\hat{\boldsymbol{d}} = \ell\right] = \ell\mathbb{P}\left[\boldsymbol{d} = \ell\right]/d, \mathbb{P}\left[\hat{\boldsymbol{k}} = \ell\right] = \ell\mathbb{P}\left[\boldsymbol{k} = \ell\right]/k \quad (\ell \geq 0). \tag{1.10}$$

Throughout the paper we let $(\boldsymbol{k}_i, \boldsymbol{d}_i, \hat{\boldsymbol{k}}_i, \hat{\boldsymbol{d}}_i)_{i\geq 1}$ denote mutually independent copies of $\boldsymbol{k}, \boldsymbol{d}, \hat{\boldsymbol{k}}, \hat{\boldsymbol{d}}$. Unless specified otherwise, all these random variables are assumed to be independent of any other sources of randomness.

We use common notation for graphs and multigraphs. For instance, for a vertex v of a multigraph G we denote by $\partial_G v$ the set of neighbors of v. More generally, for an integer $\ell \geq 1$ we let $\partial_G^\ell v$ be the set of vertices at distance precisely ℓ from v. We omit the reference to G where possible.

The proofs of the main results rely on taking a double limit where we first take the number n of columns to infinity and subsequently send an error parameter ε to zero. We use the asymptotic symbols with an index n such as $O_n(\cdot)$, $o_n(\cdot)$ to refer to the inner limit $n \to \infty$ only. Thus, for functions $f(\varepsilon, n), g(\varepsilon, n)$ we write

$$f(\varepsilon,n) = O_n(g(n,\varepsilon)) \quad \text{if pointwise for every } \varepsilon > 0, \quad \limsup_{n\to\infty}\left|\frac{f(\varepsilon,n)}{g(\varepsilon,n)}\right| < \infty,$$
$$f(\varepsilon,n) = o_n(g(n,\varepsilon)) \quad \text{if pointwise for every } \varepsilon > 0, \limsup_{n\to\infty}\left|\frac{f(\varepsilon,n)}{g(\varepsilon,n)}\right| = 0.$$

For example, $1/(\varepsilon n) = o_n(1)$. Additionally, we will use the symbols $O_{\varepsilon,n}$, $o_{\varepsilon,n}$, etc. to refer to the double limit $\varepsilon \to 0$ after $n \to \infty$. Thus,

$$f(\varepsilon,n)=O_{\varepsilon,n}(g(\varepsilon,n)) \quad \text{if} \quad \limsup_{\varepsilon\to 0}\limsup_{n\to\infty}\left|\frac{f(\varepsilon,n)}{g(\varepsilon,n)}\right|<\infty,$$
$$f(\varepsilon,n)=o_{\varepsilon,n}(g(\varepsilon,n)) \quad \text{if} \quad \limsup_{\varepsilon\to 0}\limsup_{n\to\infty}\left|\frac{f(\varepsilon,n)}{g(\varepsilon,n)}\right|=0.$$

For instance, $\varepsilon+1/(\varepsilon n)=o_{\varepsilon,n}(1)$.

Finally, we need the following basic lemma on sums of independent random variables.

Lemma 1.11. *Let $r>2$, $\delta>0$ and suppose that $(\lambda_i)_{i\geq 1}$ are independent copies of a random variable $\lambda\geq 0$ with $\mathbb{E}[\lambda^r]<\infty$. Further, let $s=\Theta_n(n)$. Then $\mathbb{P}\left[\left|\sum_{i=1}^{s}(\lambda_i-\mathbb{E}[\lambda])\right|>\delta n\right]=o_n(1/n)$.*

For the sake of completeness the proof of Lemma 1.11 is included in the appendix.

2 | OVERVIEW

We survey the proof of Theorem 1.1 and subsequently compare these techniques with those employed in prior work. The main contribution of the paper is the "≥"-part of (1.3), that is, the lower bound on the rank. We prove this lower bound via a technique inspired by the physicists' cavity method. The scaffolding of the proof is provided by a coupling argument reminiscent of a proof strategy known in mathematical physics jargon under the name "Aizenman-Sims-Starr scheme" [2] or "cavity ansatz" [31]:

> To calculate the mean of a random variable X_n on a random system of size n in the limit $n\to\infty$, calculate the difference $\mathbb{E}[X_{n+1}]-\mathbb{E}[X_n]$ upon going to a system of size $n+1$. Perform this calculation by coupling the systems of sizes n and $n+1$ such that the latter results from the former by adding only a bounded number of elements.

We will apply this approach to $X_n=\operatorname{nul}\boldsymbol{A}_n$. The coupling will be such that X_{n+1} is the nullity of a random matrix obtained from $\boldsymbol{A}_n$ obtained by adding a few rows and columns. Thus, we need to calculate the ensuing change in nullity upon adding to a matrix several rows/columns whose number is random and bounded in expectation.

In general, such a calculation hardly seems possible. To carry it out we would need to understand the linear dependencies among the coordinates where the new rows sport nonzero entries, an exceedingly complicated task. Two facts deliver us from this complexity. First, the positions of the nonzero entries of the new rows are (somewhat) random. Second, we develop a random perturbation, applicable to any matrix, that diminishes the number of short linear relations (Proposition 2.4 below). To be precise, we will conclude that by applying the perturbation, for any fixed ℓ the probability that a set of ℓ coordinates forms a proper relation in the sense of Definition 2.1 below can be made negligibly small without substantially altering the nullity. In effect, the probability that there will be linear dependencies among the positions of the nonzero entries of the new rows will turn out to be negligible. Since this perturbation argument is the linchpin of the entire proof, this is what we shall begin with. Subsequently we will explain how this general perturbation renders the desired lower bound on the rank.

2.1 | Short linear relations

Define the *support* of a vector $\xi\in\mathbb{F}^U$ as $\operatorname{supp}(\xi)=\{i\in U:\xi_i\neq 0\}$.

Definition 2.1. Let A be an $m \times n$-matrix over a field $\mathbb{F}$.

- A set $\emptyset \neq I \subseteq [n]$ is a relation of A if there exists a row vector $y \in \mathbb{F}^{1\times m}$ such that $\emptyset \neq \mathrm{supp}(yA) \subseteq I$.
- If $I = \{i\}$ is a relation of A, then we call i frozen in A. Let $\mathfrak{F}(A)$ be the set of all frozen $i \in [n]$.
- A set $I \subseteq [n]$ is a proper relation of A if $I \setminus \mathfrak{F}(A)$ is a relation of A.
- For $\delta > 0$, $\ell \geq 1$ we say that A is (δ, ℓ)-free if there are no more than δn^{ℓ} proper relations $I \subseteq [n]$ of size $|I| = \ell$.

Thus, if $I \subseteq [n]$ is a relation of A, then by adding up suitable multiples of the rows of the homogeneous linear system $Ax = 0$ we can infer a nontrivial linear relation involving the variables $(x_i)_{i\in I}$ only. In the simplest case the set $I = \{i\}$ may be a singleton. Then the equation $x_i = 0$ is implicit in $Ax = 0$ and we call coordinate i frozen. In particular, i is frozen if A contains a row whose only nonzero entry appears in column i. However, this is not the only possibility. For instance, in the following $\mathbb{F}_2$-matrix variable x_1 is frozen because the sum of all three rows equals (1 0 0):

$$\begin{pmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}. \tag{2.1}$$

In effect, for any vector ξ in the kernel of (2.1) we have

$$0 = (1\ 1\ 1)\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = (1\ 1\ 1)\left[\begin{pmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}\xi\right] = \left[(1\ 1\ 1)\begin{pmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}\right]\xi = (1\ 0\ 0)\xi = \xi_1. \tag{2.2}$$

Generally, a linear number $\Omega(n)$ of rows may have to collude to cause freezing. Moreover, although the proof is just a bit of routine linear algebra, it is worthwhile including the following characterization of frozen coordinates.

Fact 2.2. *A coordinate i is frozen in the matrix A iff $\xi_i = 0$ for all $\xi \in \ker A$.*

Proof. Let A be an $m\times n$ matrix over an arbitrary field. The calculation from (2.2) readily generalizes to arbitrary matrices and implies that $\xi_i = 0$ for any frozen coordinate $i \in [n]$ and any $\xi \in \ker A$.

Conversely, assume that for coordinate $i \in [n]$ we have $\xi_i = 0$ for all $\xi \in \ker A$. Let $e^{(i)} \in \mathbb{F}^{1\times n}$ be the vector whose ith coordinate equals one and whose other coordinates are equal to zero. Moreover, obtain A^+ from A by adding $e^{(i)}$ as an extra row. Because $\xi_i = 0$ for all $\xi \in \ker A$ we have $\ker A^+ = \ker A$. Therefore, $\mathrm{rk}\, A = \mathrm{rk} A^+$ and thus $e^{(i)}$ is a linear combination of the rows of A. Hence, $i \in \mathfrak{F}(A)$. ∎

Furthermore, excluding frozen coordinates, a proper relation I of A renders a nontrivial linear relation among at least two of the variables $(x_i)_{i\in I}$. Finally, A is (δ, ℓ)-free if only few ℓ-subsets $I \subseteq [n]$ are proper relations.

We proceed to put forward a small random perturbation that will mostly rid a given matrix of short proper relations, an observation that we expect to be of independent interest.

Definition 2.3. Let A be an $m \times n$ matrix and let $\theta \geq 0$ be an integer. Let $\boldsymbol{i}_1, \boldsymbol{i}_2, \ldots, \boldsymbol{i}_\theta \in [n]$ be uniformly random and mutually independent column indices. Then the matrix $A[\theta]$ is obtained by

adding θ new rows to A such that for each $j \in [\theta]$ the jth new row has precisely one nonzero entry, namely a one in the $\boldsymbol{i}_j$th column.

In other words, in $A[\theta]$ we expressly peg θ randomly chosen variables $x_{i_1}, \ldots, x_{i_\theta}$ of the linear system $Ax = 0$ to zero. The proof of the following proposition is based on a blend of algebraic and probabilistic ideas.

Proposition 2.4. *For any $\delta > 0$, $\ell > 0$ there exists $\mathscr{T} = \mathscr{T}(\delta, \ell) > 0$ such that for any matrix A over any field $\mathbb{F}$ the following is true. With $\boldsymbol{\theta} \in [\mathscr{T}]$ chosen uniformly at random we have*

$$\mathbb{P}[A[\boldsymbol{\theta}] \text{ is } (\delta, \ell) - \text{free}] > 1 - \delta. \tag{2.3}$$

The key feature of Proposition 2.4 is that the maximum number $\mathscr{T}$ of variables that get pegged to zero does not depend on the matrix A or its size but on δ and ℓ only. Moreover, since adding a single row can change the nullity by at most one, we obtain $|\mathrm{nul}(A) - \mathrm{nul}\, A[\boldsymbol{\theta}]| \leq \mathscr{T}$. Hence, while eliminating short proper relations, the perturbation does not shift the nullity significantly. Proposition 2.4 is a sweeping generalisation of a probabilistic result from [7], where the perturbation from Definition 2.3 was applied to matrices over finite fields to diminish stochastic dependencies among entries of randomly chosen vectors in the kernel. That argument, in turn, was inspired by ideas from information theory [17, 48, 51]. We will come back to this in Section 2.3.

We will incorporate the perturbation from Proposition 2.4 into the Aizenman–Sims–Starr coupling argument, which reduces the rank calculation to studying the impact of a few additional rows and columns on the rank. The following lemma, whose proof consists of a few lines of linear algebra, shows how the impact of such operations can be tracked in the absence of proper relations. Specifically, the lemma shows that all we need to know about the matrix A to which we add rows/columns is the set $\mathfrak{F}(A)$ of frozen variables.

Lemma 2.5. *Let A, B, C be matrices of size $m \times n$, $m' \times n$ and $m' \times n'$, respectively, and let $I \subseteq [n]$ be the set of all indices of nonzero columns of B. Moreover, obtain B_* from B by replacing for each $i \in I \cap \mathfrak{F}(A)$ the ith column of B by zero. Unless I is a proper relation of A we have*

$$\mathrm{nul}\begin{pmatrix} A & 0 \\ B & C \end{pmatrix} - \mathrm{nul}\, A = n' - \mathrm{rk}(B_*\ C). \tag{2.4}$$

Observe that the quantity on the l.h.s. of (2.4) (and thus the one on the r.h.s. as well) may be either positive or negative, depending on A, B, C.

To put Proposition 2.4 and Lemma 2.5 to work, we need to explain the construction of the telescoping series of random variables upon which the Aizenman–Sims–Starr argument is based. That is our next step.

2.2 | The Aizenman–Sims–Starr scheme

In order to derive the desired lower bound on the rank we need to bound the nullity of $\boldsymbol{A}_n$ from above. In line with the Aizenman–Sims–Starr scheme [2, 49], a first stab at this problem might be to write a telescoping sum

$$\limsup_{n\to\infty} \frac{1}{n}\mathbb{E}[\mathrm{nul}(\boldsymbol{A}_n)] = \limsup_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N-1} \mathbb{E}[\mathrm{nul}(\boldsymbol{A}_{n+1})] - \mathbb{E}[\mathrm{nul}(\boldsymbol{A}_n)]. \tag{2.5}$$

Providing that $\mathbb{E}[\mathrm{nul}(\boldsymbol{A}_{n+1})] - \mathbb{E}[\mathrm{nul}(\boldsymbol{A}_n)]$ is bounded, the lim sup of the sequence of summands exists. In this case, due to the normalizing factor $1/N$ on the r.h.s. of (2.5), we obtain

$$\limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N-1}\mathbb{E}[\mathrm{nul}(\boldsymbol{A}_{n+1})] - \mathbb{E}[\mathrm{nul}(\boldsymbol{A}_n)] \le \limsup_{n\to\infty}\mathbb{E}[\mathrm{nul}(\boldsymbol{A}_{n+1})] - \mathbb{E}[\mathrm{nul}(\boldsymbol{A}_n)]. \tag{2.6}$$

Hence, combining (2.5) and (2.6), we obtain the bound

$$\limsup_{n\to\infty}\frac{1}{n}\mathbb{E}[\mathrm{nul}(\boldsymbol{A}_n)] \le \limsup_{n\to\infty}\mathbb{E}[\mathrm{nul}(\boldsymbol{A}_{n+1})] - \mathbb{E}[\mathrm{nul}(\boldsymbol{A}_n)].$$

To obtain an explicit estimate, we should thus attempt to couple $\boldsymbol{A}_{n+1}$ and $\boldsymbol{A}_n$ so that we can write a single expectation

$$\mathbb{E}[\mathrm{nul}(\boldsymbol{A}_{n+1})] - \mathbb{E}[\mathrm{nul}(\boldsymbol{A}_n)] = \mathbb{E}\left[\mathrm{nul}(\boldsymbol{A}_{n+1}) - \mathrm{nul}(\boldsymbol{A}_n)\right]. \tag{2.7}$$

Ideally, to bring the tools from Section 2.1 to bear, under this coupling $\boldsymbol{A}_{n+1}$ should be obtained from $\boldsymbol{A}_n$ by adding one column and a few rows.

Unfortunately, this direct approach flounders for obvious reasons. For instance, depending on the distributions $\boldsymbol{d}, \boldsymbol{k}$, due to divisibility issues $\boldsymbol{A}_{n+1}$ may not even be defined for all n.[5] To deal with this issue we introduce a more malleable version of the random matrix model, without significantly altering the rank. Specifically, we introduce a parameter $\varepsilon > 0$, for which we choose a large enough $\mathscr{T} = \mathscr{T}(\varepsilon) > 0$. Then for integers $n \ge \mathscr{T}$ we construct a random matrix $\boldsymbol{A}_{\varepsilon,n}$ as follows. Like in Section 1.2 let $\chi : [0,1]^2 \to \mathbb{F}^*$ be a measurable map and let $(\boldsymbol{\zeta}_i, \boldsymbol{\xi}_i)_{i\ge1}$ be uniformly distributed $[0,1]$-valued random variables. Further, let

$$\boldsymbol{m}_{\varepsilon,n} \sim \mathrm{Po}((1-\varepsilon)dn/k).$$

Additionally, choose $\boldsymbol{\theta} \in [\mathscr{T}]$ uniformly at random and, as before, let $(\boldsymbol{d}_i)_{i\ge1}$, $(\boldsymbol{k}_i)_{i\ge1}$ be copies of $\boldsymbol{d}$, $\boldsymbol{k}$. All of these random variables are mutually independent. Further, let $\boldsymbol{\Gamma}_{\varepsilon,n}$ be a uniformly random maximal matching of the complete bipartite graph with vertex classes

$$\bigcup_{i=1}^{\boldsymbol{m}_{\varepsilon,n}}\{a_i\}\times[\boldsymbol{k}_i] \quad \text{and} \quad \bigcup_{j=1}^{n}\{x_j\}\times[\boldsymbol{d}_j].$$

As in the well known configuration model of random graphs, we think of $\{a_i\}\times[\boldsymbol{k}_i]$ as a set of clones of a_i and of $\{x_j\}\times[\boldsymbol{d}_j]$ as a set of clones of x_j. We obtain a random Tanner graph $\boldsymbol{G}_{\varepsilon,n}$ with variable nodes $x_1, \ldots, x_n$ and check nodes $a_1, \ldots, a_{\boldsymbol{m}_{\varepsilon,n}}, p_1, \ldots, p_{\boldsymbol{\theta}}$ by inserting an edge between a_i and x_j for each matching edge that joins the sets $\{a_i\}\times[\boldsymbol{k}_i]$ and $\{x_j\}\times[\boldsymbol{d}_j]$. Additionally, check node p_i is adjacent to x_i for each $i \in [\boldsymbol{\theta}]$. To be clear, we do not need to set aside any unmatched variable clones as partners of the p_i. We simply add the x_i-p_i-edges on top of the configuration model. Since ultimately $\mathscr{T}$ will be chosen to be of order $o(n)$, the number of these additional edges is relatively small.

Since there may be several edges joining clones of the same variable and check node, $\boldsymbol{G}_{\varepsilon,n}$ may be a multigraph. Finally, we construct a random matrix $\boldsymbol{A}_{\varepsilon,n}$ whose rows are indexed by the check

[5]For instance, suppose that $\boldsymbol{d} = 3$ and $\boldsymbol{k} = 4$ deterministically. Then (1.1) boils down to $4\boldsymbol{m} = 3n$, and thus $\boldsymbol{A}_n$ is well-defined only if n is divisible by four.

nodes $a_1, \ldots, a_{\boldsymbol{m}_{\varepsilon,n}}$ and whose columns are indexed by $x_1, \ldots, x_n$ such that the nonzero entries of $\boldsymbol{A}_{\varepsilon,n}$ represent the edges of the matching $\boldsymbol{\Gamma}_{\varepsilon,n}$. Specifically, the matrix entries read

$$(\boldsymbol{A}_{\varepsilon,n})_{p_i,x_j} = \mathbf{1}\{i=j\} \qquad (i \in [\boldsymbol{\theta}], j \in [n]),$$
$$(\boldsymbol{A}_{\varepsilon,n})_{a_i,x_j} = \chi_{\zeta_i,\xi_j} \sum_{s=1}^{k_i} \sum_{t=1}^{\boldsymbol{d}_j} \mathbf{1}\left\{\{(a_i,s),(x_j,t)\} \in \boldsymbol{\Gamma}_{\varepsilon,n}\right\} \quad (i \in [\boldsymbol{m}_{\varepsilon,n}], j \in [n]).$$

Morally, $\boldsymbol{A}_{\varepsilon,n}$ mimics the matrix obtained from the original model $\boldsymbol{A}_n$ by deleting every row with probability ε independently (which, of course, would be unworkable because still the model is not generally defined for all n). Furthermore, the purpose of the check nodes $p_1, \ldots, p_{\boldsymbol{\theta}}$ is to ensure that $\boldsymbol{A}_{\varepsilon,n}$ is (δ,ℓ)-free for a small enough $\delta = \delta(\varepsilon)$ and a large enough $\ell = \ell(\varepsilon)$. Indeed, while Proposition 2.4 requires that a *random* set of $\boldsymbol{\theta}$ variables be pegged, the checks $p_1, \ldots, p_{\boldsymbol{\theta}}$ just freeze the first $\boldsymbol{\theta}$ variables. But since the distribution of the Tanner graph $\boldsymbol{G}_{\varepsilon,n} - \{p_1, \ldots, p_{\boldsymbol{\theta}}\}$ is invariant under permutations of the variable nodes, both constructions are equivalent. The following proposition shows that going to $\boldsymbol{A}_{\varepsilon,n}$ does not shift the rank significantly.

Proposition 2.6. *For any any $0 < C < C'$ and any function $\mathcal{T} = \mathcal{T}(\varepsilon) \geq 0$ the following is true. If*

$$\limsup_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{n}\mathbb{E}[\mathrm{nul}(\boldsymbol{A}_{\varepsilon,n})] \leq C \quad \text{then} \quad \lim_{n\to\infty} \mathbb{P}\left[\mathrm{nul}(\boldsymbol{A}_n) \leq C'n\right] = 1. \tag{2.8}$$

Analogously, if

$$\liminf_{\varepsilon\to 0} \liminf_{n\to\infty} \frac{1}{n}\mathbb{E}[\mathrm{nul}(\boldsymbol{A}_{\varepsilon,n})] \geq C' \quad \text{then} \quad \lim_{n\to\infty} \mathbb{P}\left[\mathrm{nul}(\boldsymbol{A}_n) \geq Cn\right] = 1.$$

By construction, the degrees of the checks a_i and the variables x_j in $\boldsymbol{G}_{\varepsilon,n} - \{p_1, \ldots, p_{\boldsymbol{\theta}}\}$ are upper-bounded by $\boldsymbol{k}_i$ and $\boldsymbol{d}_j$, respectively. We thus refer to $\boldsymbol{k}_i$ and $\boldsymbol{d}_j$ as the *target degrees* of a_i and x_j. Indeed, since $\boldsymbol{G}_{\varepsilon,n}$ will turn out to feature few if any multi-edges and $\boldsymbol{m}_{\varepsilon,n}$ is significantly smaller than dn/k and thus

$$\mathbb{P}\left[\sum_{i=1}^{\boldsymbol{m}_{\varepsilon,n}} \boldsymbol{k}_i \leq \sum_{i=1}^{n} \boldsymbol{d}_i\right] = 1 - o_n(1),$$

most check nodes a_i have degree precisely $\boldsymbol{k}_i$ a.a.s. But we expect that about εdn variable nodes x_i will have degree less than $\boldsymbol{d}_i$. In fact, a.a.s. $\boldsymbol{\Gamma}_{\varepsilon,n}$ fails to cover about εdn "clones" from the set $\bigcup_{i=1}^{n}\{x_i\}\times[\boldsymbol{d}_i]$. Let us call such unmatched clones *cavities*.

The cavities provide the wiggle room that we need to couple $\boldsymbol{A}_{\varepsilon,n}$ and $\boldsymbol{A}_{\varepsilon,n+1}$. An instant idea might be to couple $\boldsymbol{G}_{\varepsilon,n+1}$ and $\boldsymbol{G}_{\varepsilon,n}$ such that the former is obtained by adding one variable node x_{n+1} along with $\boldsymbol{d}_{n+1}$ new adjacent check nodes. Additionally, the new checks get connected with some random cavities of $\boldsymbol{G}_{\varepsilon,n}$. In effect, the coupling takes the form

$$\mathrm{nul}\,\boldsymbol{A}_{\varepsilon,n+1} = \mathrm{nul}\begin{pmatrix}\boldsymbol{A}_{\varepsilon,n} & 0\\ \boldsymbol{B} & \boldsymbol{C}\end{pmatrix}, \tag{2.9}$$

where $\boldsymbol{B}$ has n columns and $\boldsymbol{d}_{n+1}$ rows and $\boldsymbol{C}$ is a column vector of size $\boldsymbol{d}_{n+1}$ a.a.s. But this direct attempt has a subtle flaw. Indeed, going from $\boldsymbol{A}_{\varepsilon,n}$ to $\boldsymbol{A}_{\varepsilon,n+1}$, (2.9) adds $\mathbb{E}[\boldsymbol{d}_{n+1}] = d$ rows on the average. Yet

actually we should be adding merely $\mathbb{E}[\boldsymbol{m}_{\varepsilon,n+1} - \boldsymbol{m}_{\varepsilon,n}] = (1-\varepsilon)d/k$ rows. To remedy this problem we borrow a trick from prior applications of the Aizenman–Sims–Starr scheme in combinatorics [7, 17, 18]. Namely, we set up a coupling under which both $\boldsymbol{A}_{\varepsilon,n}, \boldsymbol{A}_{\varepsilon,n+1}$ are obtained by adding a few rows/columns to a common "base matrix" $\boldsymbol{A}'$. Thus, instead of (2.9) we obtain

$$\operatorname{nul} \boldsymbol{A}_{\varepsilon,n} = \operatorname{nul} \begin{pmatrix} \boldsymbol{A}' \\ \boldsymbol{B} \end{pmatrix}, \quad \operatorname{nul} \boldsymbol{A}_{\varepsilon,n+1} = \operatorname{nul} \begin{pmatrix} \boldsymbol{A}' & 0 \\ \boldsymbol{B}' & \boldsymbol{C}' \end{pmatrix}. \tag{2.10}$$

To be precise, $\boldsymbol{C}'$ above is a column vector with an expected $(1-\varepsilon)d$ nonzero entries and $\boldsymbol{B}, \boldsymbol{B}'$ are matrices whose numbers of nonzero entries are bounded in expectation. Furthermore, the base matrix $\boldsymbol{A}'$ itself is quite similar to $\boldsymbol{A}_{\varepsilon,n}$, except that $\boldsymbol{A}'$ has a slightly smaller number of rows. In Section 5 we will present the construction in full detail and apply Proposition 2.4 and Lemma 2.5 to prove the following upper bound on the change in nullity. Recall the function Φ from (1.2) and recall that $\boldsymbol{\theta} \in [\mathscr{T}\,]$ with $\mathscr{T} = \mathscr{T}(\varepsilon)$ dependent on ε only is the number of pinned variables in the construction of $\boldsymbol{A}_{\varepsilon,n}$.

Proposition 2.7. *There exists a function $\mathscr{T} = \mathscr{T}(\varepsilon) > 0$ such that*

$$\limsup_{\varepsilon\to 0} \limsup_{n\to\infty} \mathbb{E}[\operatorname{nul}(\boldsymbol{A}_{\varepsilon,n+1})] - \mathbb{E}[\operatorname{nul}(\boldsymbol{A}_{\varepsilon,n})] \le \max_{\alpha\in[0,1]} \Phi(\alpha).$$

As an immediate consequence of Proposition 2.7 we obtain the desired upper bound on the nullity.

Corollary 2.8. *We have*

$$\limsup_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{n}\mathbb{E}[\operatorname{nul}(\boldsymbol{A}_{\varepsilon,n})] \le \max_{\alpha\in[0,1]} \Phi(\alpha).$$

Proof. Proposition 2.7 yields

$$\frac{1}{n}\mathbb{E}[\operatorname{nul}(\boldsymbol{A}_{\varepsilon,n})] = \frac{1}{n}\left[\mathbb{E}[\operatorname{nul}(\boldsymbol{A}_{\varepsilon,1})] + \sum_{N=1}^{n-1}\left(\mathbb{E}[\operatorname{nul}(\boldsymbol{A}_{\varepsilon,N+1})] - \mathbb{E}[\operatorname{nul}(\boldsymbol{A}_{\varepsilon,N})]\right)\right] \le \max_{\alpha\in[0,1]} \Phi(\alpha) + o_{\varepsilon,n}(1),$$

as claimed. ∎

Proof of Theorem 1.1. The desired lower bound on the rank of $\boldsymbol{A}_n$ is an immediate consequence of Proposition 2.6 and Corollary 2.8. ∎

2.3 | Discussion

Before delving into the technical details of the proofs of the various propositions, we compare the proof strategy and the results with previous work. We begin with a discussion of related work on the rank problem. Roughly speaking, prior work on the rank of random matrices relies on separate strands of techniques, depending on whether the average number of nonzero entries per row/column is bounded or unbounded. Subsequently we discuss the physicists' (nonrigorous) cavity method and explain how it led to an erroneous prediction.

2.3.1 | Dense matrices

The difficulty of the rank problem for dense random matrices strongly depends on the distribution of the matrix entries. For instance, a square matrix with independent Gaussian entries in each row has full

rank with probability one simply because the submanifold of singular matrices has Lebesgue measure zero. By contrast, the case of matrices with independent uniform ± 1 entries is more subtle. Komlós [36] proved that such matrices are regular a.a.s. Vu [55] subsequently presented a simpler proof, based on collision probabilities and Erdős' Littlewood-Offord inequality. An intriguing conjecture, which has inspired a distinguished line of research [33, 34, 44, 53–54], asserts that the dominant reason for a random ± 1-matrix being singular is the existence of a pair of identical rows or columns.

Interesting enough, the limiting probability that a dense square matrix with entries drawn uniformly from a finite field is singular lies strictly between zero and one. Kovalenko and Levitskaya [37, 38, 42, 43] obtained a precise formula for the distribution of the rank of dense random matrices with independent entries over finite fields via the method of moments. For more recent improvements see [27] and the references therein.

A further line of work deals with random $m \times n$ matrices in which the number of nonzero entries per row diverges in the limit of large n but is of order $o_n(n)$. Relating the permanent and the determinant, Balakin [9] and, using delicate moment calculations, Blömer, Karp and Welzl [12] dealt with the rank of such matrices over finite fields. Moreover, using expansion arguments, Costello and Vu [21, 22] studied the real rank of random symmetric matrices of a similar density. They find that such matrices essentially have full rank a.a.s., apart from a small defect based on local phenomena. In the words of [22], "dependency [comes] from small configurations".

2.3.2 | Sparse matrices

Matters are quite different in the sparse case where the average number of nonzero entries per row is bounded. In fact, as we will discover in due course the formula from Theorem 1.1 is driven by "dependency coming from large configurations," that is, by minimally linearly dependent sets of unbounded size.

The first major contribution dedicated to sparse matrices was a paper by Dubois and Mandler [24] on the random 3-XORSAT problem. Translated into random matrices, this problem asks for what ratios m/n a random $m \times n$-matrix over $\mathbb{F}_2$ with precisely three one-entries in each row has full rank (i.e., equal to $m \wedge n$) a.a.s. Thus, the random matrix model is just the one from Example 1.8 with $k = 3$. Dubois and Mandler pinpointed the precise full row rank threshold $m/n \approx 2.75$. The proof relies on the first moment method applied to $|\ker \boldsymbol{A}|$, which boils down to a one-dimensional calculus problem. Matters get more complicated when one considers a greater number $k > 3$ of nonzero entries per row. This more general problem, known as random k-XORSAT, was solved independently by Dietzfelbinger et al. [23] and by Pittel and Sorkin [50] via technically demanding moment calculations. Unfortunately, considering fields $\mathbb{F}_q$ with $q > 2$ complicates the moment calculation even further. Yet undertaking a computer-assisted tour-de-force Falke and Goerdt [29] managed to extend the method to $\mathbb{F}_3$. However, extending this strategy to infinite fields is a nonstarter as $|\ker \boldsymbol{A}|$ may be infinite.

In a previous paper Ayre, Coja-Oghlan, Gao and Müller [7] applied the Aizenman–Sims–Starr scheme to the study of sparse random matrices with precisely k nonzero entries per row as in Example 1.8, over finite fields. The present paper goes beyond that earlier contribution in two crucial ways. First, we develop a far more delicate coupling scheme that accommodates general degree distributions $\boldsymbol{d}, \boldsymbol{k}$ rather than just the Poisson-constant degrees from Example 1.8, including degree sequences for which the 2-core bound fails to be tight (in contrast to Example 1.8). Apart from rendering a proof of Lelarge's conjecture, we expect that this more general coupling scheme will find further uses in the theory of random factor graphs; for instance, it seems applicable to generalizations of the models from [17].

Second, the rank calculation in [7] is based on a probabilistic view that does not extend to infinite fields. Indeed, the proof there is based on a close study of a uniformly random element $\boldsymbol{\sigma}$ of the kernel of the random matrix $\boldsymbol{A}$. Specifically, [7, lemma 3.1] analyzes the impact of the perturbation from Definition 2.3 on a matrix $A \in \mathbb{F}^{m\times n}$ for a finite field $\mathbb{F}$. With $\boldsymbol{\sigma} = (\boldsymbol{\sigma}_1, \ldots, \boldsymbol{\sigma}_n) \in \ker(A)$ a uniformly random element of the kernel, the lemma shows that for a large enough $\mathscr{T} = \mathscr{T}(\delta, \mathbb{F}) > 0$ and a uniformly random $0 \le \boldsymbol{\theta} \le \mathscr{T}$,

$$\sum_{\substack{1\le i<j\le n\\ \omega,\omega'\in\mathbb{F}}} \mathbb{E}\left|\mathbb{P}\left[\boldsymbol{\sigma}_i = \omega, \boldsymbol{\sigma}_j = \omega' | A[\boldsymbol{\theta}]\right] - \mathbb{P}\left[\boldsymbol{\sigma}_i = \omega | A[\boldsymbol{\theta}]\right]\cdot\mathbb{P}\left[\boldsymbol{\sigma}_j = \omega' | A[\boldsymbol{\theta}]\right]\right| < \delta n^2, \tag{2.11}$$

As in Proposition 2.4, the necessary value of $\mathscr{T}$ is independent of n, m, and A. Thus, the random perturbation renders the vector entries $(\boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j)$ nearly stochastically independent, for most i, j. Thanks to general results from [10], (2.11) extends from pairwise independence to ℓ-wise independence, albeit with a weaker error bound δ. The result [7, lemma 3.1] was inspired by general statements about probability measures on discrete cubes from [17, 48, 51].

Inherently, this stochastic approach does not generalize to infinite fields, where, for starters, it does not even make sense to speak of a uniformly random element of the kernel. That is why here we replace the stochastic approach from the earlier paper by the more versatile algebraic approach summarized in Proposition 2.4, which are applicable to any field—say, the reals, the field $\mathbb{Q}_p$ of p-adic numbers, the algebraic closure of a finite field or a structure as complex as a function field. Instead of showing stochastic independence, Proposition 2.4 renders linear independence amongst most bounded-size subsets of coordinates. Apart from being more general, this algebraic viewpoint allows for a cleaner, more direct proof of the rank formula. Additionally, on finite fields the stochastic independence (2.11) follows from the linear independence provided by Proposition 2.4, with a significantly improved bound on $\mathscr{T}(\delta)$. We work this out in detail in Appendix B.

The single prior contribution on the rational rank of sparse random matrices is due to Bordenave, Lelarge and Salez [13], who computed the rational rank of the (symmetric) adjacency matrix of a random graph with a given vertex degree distribution. The proof is based on local weak convergence and the "objective method" [5]. An intriguing question for future research is to extend the techniques from the present paper to symmetric random matrices.

2.3.3 | The cavity method (and its caveats)

On the basis of the cavity method, an analytic but nonrigorous technique inspired by the statistical mechanics of disordered systems, it had been predicted erroneously that over finite fields the 2-core bound (1.7) on the rank of $\boldsymbol{A}$ is universally tight for general degree distributions $\boldsymbol{d}, \boldsymbol{k}$ [3, 46]. Where did the cavity method go astray?

The method comes in two instalments, the simpler *replica symmetric ansatz* and the more elaborate *one-step replica symmetry breaking ansatz* ("1RSB"). The former predicts that the rank of $\boldsymbol{A}$ over a finite field $\mathbb{F}_q$ converges in probability to the solution of an optimization problem on an infinite-dimensional space of probability measures. To be precise, let $\mathscr{P}(\mathbb{F}_q)$ be the space of probability measures on $\mathbb{F}_q$. Identify this space with the standard simplex in $\mathbb{R}^q$. Further, let $\mathscr{P}^2(\mathbb{F}_q)$ be the space of all probability measures on $\mathscr{P}(\mathbb{F}_q)$. Given $\pi \in \mathscr{P}^2(\mathbb{F}_q)$ let $(\boldsymbol{\mu}^{\pi}_{i,j})_{i,j\ge1}$ be a sequence of independent samples from π. Recalling (1.10), the *Bethe free entropy* is defined by

$$\mathscr{B}(\pi) = \mathbb{E}\left[\log_q \sum_{\sigma_1\in\mathbb{F}_q}\prod_{i=1}^{d}\sum_{\sigma_2,\ldots,\sigma_{\hat{k}_i}\in\mathbb{F}_q}\mathbf{1}\left\{\sum_{j=1}^{\hat{k}_i}\sigma_j\boldsymbol{\chi}_{i,j}=0\right\}\prod_{j=2}^{\hat{k}_i}\boldsymbol{\mu}^{\pi}_{i,j}(\sigma_j)\right]$$

$$-\frac{d}{k}\mathbb{E}\left[(k-1)\log_q \sum_{\sigma_1,\ldots,\sigma_k\in\mathbb{F}_q} \mathbf{1}\left\{\sum_{i=1}^{k}\sigma_i\chi_{1,i}=0\right\}\prod_{i=1}^{k}\mu_{1,i}^{\pi}(\sigma_i)\right].$$

(cf. [73, chapter 14]).

The replica symmetric ansatz predicts that

$$\lim_{n\to\infty}\frac{1}{n}\mathrm{nul}\,\boldsymbol{A}=\sup_{\pi\in\mathscr{P}^2(\mathbb{F}_q)}\mathscr{B}(\pi)\quad\text{in probability.}\tag{2.12}$$

For a detailed (heuristic) derivation of the Bethe free entropy and the prediction (2.12) we refer to [3]. But let us briefly comment on the intended semantics of π. Consider the Tanner graph $\boldsymbol{G}$ representing $\boldsymbol{A}$. Suppose that variable node x_i and check node a_j are adjacent. Then for $\sigma\in\mathbb{F}_q$ we define the *Belief Propagation message* $\mu_{\boldsymbol{A},x_j\to a_i}(\sigma)$ from x_j to a_i as follows. Obtain $\boldsymbol{A}_{x_j\to a_i}$ from $\boldsymbol{A}$ by changing the ijth matrix entry to zero; this corresponds to deleting the x_j-a_i-edge from the Tanner graph. Then $\mu_{\boldsymbol{A},x_j\to a_i}(\sigma)$ is the probability that in a uniformly random vector $\boldsymbol{\sigma}\in\ker\boldsymbol{A}_{x_j\to a_i}$ we have $\boldsymbol{\sigma}_j=\sigma$. Further, define $\pi_{\boldsymbol{A}}$ as the empirical distribution of the $\mu_{\boldsymbol{A},x_j\to a_i}$ over the edges of the Tanner graph:

$$\pi_{\boldsymbol{A}}=\frac{1}{\sum_{i=1}^{n}\boldsymbol{d}_i}\sum_{j=1}^{n}\sum_{i=1}^{\boldsymbol{m}}\mathbf{1}\{\boldsymbol{A}_{ij}\neq 0\}\delta_{\mu_{\boldsymbol{A},x_j\to a_i}}\in\mathscr{P}^2(\mathbb{F}_q).$$

Then the replica symmetric ansatz predicts that $\pi_{\boldsymbol{A}}$ is asymptotically a maximiser of the Bethe free energy, that is, $\sup_{\pi\in\mathscr{P}^2(\mathbb{F}_q)}\mathscr{B}(\pi)=\mathscr{B}(\pi_{\boldsymbol{A}})+o_n(1)$ a.a.s. Thus, the maximizer π in (2.12) is deemed to encode the Belief Propagation messages on the edges of the Tanner graph of $\boldsymbol{A}$.

A bit of linear algebra that seems to have gone unnoticed in the physics literature reveals that the messages actually have a very special form [7, lemma 2.3]. Namely, any message $\mu_{\boldsymbol{A},x_j\to a_i}$ is either the uniform distribution $q^{-1}\mathbf{1}$ on $\mathbb{F}_q$ or the atom δ_0 on 0. In effect, the rank should come out as the Bethe free entropy $\mathscr{B}(\pi_\alpha)$ of a convex combination

$$\pi_\alpha=\alpha\delta_{\delta_0}+(1-\alpha)\delta_{q^{-1}\mathbf{1}}\quad(\alpha\in[0,1]).\tag{2.13}$$

In fact, a simple calculation yields $\Phi(\alpha)=\mathscr{B}(\pi_\alpha)$ for all $\alpha\in[0,1]$. Thus, Theorem 1.1 shows that

$$\lim_{n\to\infty}\frac{\mathrm{rk}\,\boldsymbol{A}}{n}=1-\sup_{\alpha\in[0,1]}\mathscr{B}(\pi_\alpha)\quad\text{in probability,}$$

vindicating the cavity method to an extent. However, we do not know whether the Bethe free entropy possesses other spurious maximizers $\pi\in\mathscr{P}^2(\mathbb{F}_q)$ with $\mathscr{B}(\pi)>\sup_{\alpha\in[0,1]}\mathscr{B}(\pi_\alpha)$.

Alamino and Saad [3] tackled the optimisation problem (2.12) by means of a numerical heuristic called population dynamics, without noticing the restriction to $(\pi_\alpha)_{\alpha\in[0,1]}$. In all the examples that they studied they found that $\pi\in\{\pi_0,\pi_\rho\}$, with ρ from (1.5); in fact, all their examples fall within the purview of Theorem 1.4. [6] This led Alamino and Saad to conjecture that the maximiser π is generally of this form, although they cautioned that further evidence seems necessary. Example 1.9 and [41]

[6] Strictly speaking, Alamino and Saad, who worked numerically with n in the hundreds, reported $\pi\in\{\pi_0,\pi_1\}$. Indeed, $\rho\in\{0,1\}$ in the first class of examples that they studied, but not in the other two. For instance, in their example (3) the actual value of ρ is either 0 or a number strictly smaller than one, although $\rho>0.97$ whenever $\Phi(\rho)>\Phi(0)$.

provide counterexamples. The more sophisticated 1RSB cavity method is presented in [46, chapter 19], where an exercise asks the reader to verify that the 2-core bound is tight (over finite fields). While Theorem 1.4 gives sufficient conditions for this to be correct, the aforementioned counterexamples apply.

2.4 | Organization

We proceed to prove Proposition 2.4, the "key lemma" upon which the proof of Theorem 1.1 rests, in Section 3. Subsequently in Section 4 we use concentration inequalities and the local limit theorem for sums of independent random variables to prove Proposition 2.6. Additionally, Section 4 contains Proposition 1.10, which shows that the random matrix model (1.1) is well defined, a standard argument that we include for the sake of completeness. Dealing with the full details of the coupling scheme, Section 5 contains the proof of Proposition 2.7. Further, Section 6 deals with the proof of Theorem 1.2 and in Section 7 we prove Theorem 1.4. For the sake of completeness a proof of Lemma 1.11 is included in Appendix A. Moreover, in Appendix B we elaborate on the relation between the algebraic perturbation from Proposition 2.4 and the stochastic version from [7]. Finally, Appendix C contains a self-contained proof of the upper bound on the rank for Theorem 1.1 via the interpolation method from mathematical physics.

3 | LINEAR RELATIONS: PROOF OF PROPOSITION 2.4

In this section we prove Proposition 2.4 and Lemma 2.5. The somewhat delicate proof of the former is based on a blend of probabilistic and algebraic arguments. The proof of the latter is purely algebraic and fairly elementary.

3.1 | Proof of Proposition 2.4

Observe that Proposition 2.4 is not an asymptotic statement to the extent that we need to exhibit a function $\mathcal{T} = \mathcal{T}(\delta, \ell)$ such that (2.3) holds for all matrices A (ultimately in (3.12) we will see that $\mathcal{T}(\delta, \ell)$ scaling as ℓ^3/δ^4 does the trick). Nevertheless, letting n denote the number of columns of A, we may safely assume that $n > n_0 = n_0(\delta, \ell)$ for any specific n_0 that depends on δ, ℓ only. Indeed, to deal with $n \le n_0$ for any fixed value n_0 we could just pick $\mathcal{T} \ge \mathcal{T}_0(\delta, \ell)$ for a large enough $\mathcal{T}_0(\delta, \ell)$ so that with probability at least $1 - \varepsilon$ we have $\{\boldsymbol{i}_1, \ldots, \boldsymbol{i}_{\mathcal{T}}\} = [n]$. Note that we do not need to worry about the possibility that $\mathcal{T} > n$ because the $\boldsymbol{i}_j$ are drawn with replacement. Further, if $\{\boldsymbol{i}_1, \ldots, \boldsymbol{i}_{\mathcal{T}}\} = [n]$, then a glimpse at Definition 2.1 shows that *all* coordinates are frozen. Therefore, $A[\boldsymbol{\theta}]$ is (δ, ℓ)-free. Hence, from now on we assume that $n \ge n_0 = n_0(\delta, \ell)$ for a sufficiently large n_0.

Given any matrix M we define a *minimal h-relation of M* as a relation I of M of size $|I| = h$ that does not contain a proper subset that is a relation of M. Let $\mathscr{R}_h(M)$ be the set of all minimal h-relations of M and set $R_h(M) = |\mathscr{R}_h(M)|$. Thus, $R_1(M) = |\mathfrak{F}(M)|$ is just the number of frozen variables of M. Additionally, let $\mathscr{R}_{\le h}(M) = \bigcup_{1 \le i \le h} \mathscr{R}_i(M)$ and $R_{\le h}(M) = |\mathscr{R}_{\le h}(M)|$. Let $\boldsymbol{i}_1, \boldsymbol{i}_2, \boldsymbol{i}_3, \ldots \in [n]$ be uniformly distributed independent random variables.

The proof of Proposition 2.4 is based on a potential function argument. To get started we observe that

$$\mathscr{R}_1(A[t]) \subseteq \mathscr{R}_1(A[t+1]) \quad \text{for all } t \ge 0. \tag{3.1}$$

This inequality implies that the random variable

$$\Delta_t = \frac{\mathbb{E}[R_1(A[t+\ell])|A[t]] - R_1(A[t])}{n}$$

is nonnegative. The random variable Δ_t gauges the increase in frozen variables upon addition of ℓ more rows that expressly freeze specific variables. Thus, "big" values of Δ_t, say $\Delta_t = \Omega_n(1)$, witness a kind of instability as pegging a few variables to zero entails that another $\Omega_n(n)$ variables get frozen to zero due to implicit linear relations. We will exploit the observation that, since $\Delta_t \in [0,1]$ and $\mathbb{E}[\Delta_t]$ is monotonically increasing in t, such instabilities cannot occur for many t. Thus, the expectation $\mathbb{E}[\Delta_\theta]$ will serve as our potential. A similar potential was used in [7] to study stochastic dependencies in the case of finite fields $\mathbb{F}$. But in the present more general context the analysis of the potential is significantly more subtle. The following lemma puts a lid on the potential.

Lemma 3.1. *We have* $\mathbb{E}[\Delta_\theta] \le \ell/\mathscr{T}$.

Proof. For any $r \in \{0,1,\ldots,\ell-1\}$ we have

$$\sum_{j\ge 0}\mathbb{E}[\Delta_{r+j\ell}] = \frac{1}{n}\sum_{j\ge 0}\mathbb{E}[R_1(A[r+(j+1)\ell])] - \mathbb{E}[R_1(A[r+j\ell])] \le \frac{1}{n}\lim_{j\to\infty}\mathbb{E}[R_1(A[r+j\ell])]. \quad (3.2)$$

Observe that there is no problem here taking the limit $j\to\infty$ as the coordinates $\boldsymbol{i}_j$ from Definition 2.3 are chosen independently with replacement. In the case that $j \gg n$ the likely outcome is thus that *all* coordinates of $A[r+j\ell]$ are frozen, which is why $\lim_{j\to\infty}\mathbb{E}[R_1(A[r+j\ell])] = n$. Hence, (3.2) yields

$$\sum_{j\ge 0}\mathbb{E}[\Delta_{r+j\ell}] \le 1. \quad (3.3)$$

Summing (3.3) on r, we obtain

$$\sum_{\theta\in[\mathscr{T}]}\mathbb{E}[\Delta_\theta] \le \sum_{r=0}^{\ell-1}\sum_{j\ge 0}\mathbb{E}[\Delta_{r+j\ell}] \le \ell. \quad (3.4)$$

Since $\boldsymbol{\theta} \in [\mathscr{T}]$ is chosen uniformly and independently of everything else, dividing (3.4) by $\mathscr{T}$ yields

$$\mathbb{E}[\Delta_\theta] = \frac{1}{\mathscr{T}}\sum_{\theta\in[\mathscr{T}]}\mathbb{E}[\Delta_\theta] \le \frac{1}{\mathscr{T}}\sum_{r=0}^{\ell-1}\sum_{j\ge 0}\mathbb{E}[\Delta_{r+j\ell}] \le \frac{\ell}{\mathscr{T}}, \quad (3.5)$$

as desired. ∎

Remark 3.2. Lemma 3.1 provides a bound on the mean of $\mathbb{E}[\Delta_\theta]$ for a random $\boldsymbol{\theta}$. The requirement that $\boldsymbol{\theta}$ be random stems from the fact that the proof is based on an averaging argument. It is an open question whether this random value could be replaced by a deterministic value, and whether the choice of such a deterministic value would have to depend on A.

The following lemma shows that unless $A[t]$ is (δ,ℓ)-free, there exist many minimal h-relations for some $2 \le h \le \ell$.

Lemma 3.3. *If $A[t]$ fails to be (δ,ℓ)-free then there exists $2 \le h \le \ell$ such that $R_h(A[t]) \ge \delta n^h/\ell$.*

Proof. Assume that

$$R_h(A[t]) < \delta n^h/\ell \quad \text{for all } 2 \le h \le \ell. \tag{3.6}$$

Since every proper relation I of size $|I| = \ell$ contains a minimal h-relation $J \subseteq I$ for some $2 \le h \le \ell$, (3.6) implies that $A[t]$ possesses fewer than δn^ℓ proper relations of size ℓ in total. Hence, if (3.6) holds, then $A[t]$ is (δ,ℓ)-free. ∎

As a next step we show that Δ_t is large if $A[t]$ possesses many minimal h-relations for some $2 \le h \le \ell$.

Lemma 3.4. *If $R_h(A[t]) \ge \delta n^h/\ell$ for some $2 \le h \le \ell$, then $\Delta_t \ge \delta^2/\ell^2$.*

Proof. Let $\mathscr{R}_{v,h}(A[t])$ be the set of all relations $I \in \mathscr{R}_h(A[t])$ that contain $v \in [n]$ and set $r_{v,t,h} = |\mathscr{R}_{v,h}(A[t])|$. Moreover, let $\mathscr{V}_{t,h}$ be the set of all $v \in [n]$ with $r_{v,t,h} \ge \delta h n^{h-1}/(2\ell)$. We assumed $|R_h(A[t])| \ge \delta n^h/\ell$, and every h-relation is affiliated with an h-element subset of $[n]$. Consequently,

$$\delta h n^h/\ell \le h R_h(A[t]) \le |\mathscr{V}_{t,h}| n^{h-1} + \left(n - |\mathscr{V}_{t,h}|\right) \cdot \delta h n^{h-1}/(2\ell),$$

whence

$$|\mathscr{V}_{t,h}| \ge \frac{\delta h n}{2\ell}. \tag{3.7}$$

Consider $v \in \mathscr{V}_{t,h}$ along with a minimal h-relation $I \in \mathscr{R}_{v,h}(A[t])$. If $I = \{v, \boldsymbol{i}_{t+1}, \ldots, \boldsymbol{i}_{t+h-1}\}$, that is, I comprises v and the next $h-1$ indices that get pegged, then $v \in \mathfrak{F}(A[t+h-1])$. Indeed, since I is a minimal h-relation of $A[t]$ there is a row vector y such that $\mathrm{supp}(yA[t]) = I$. Hence, if $I \setminus \{v\} = \{\boldsymbol{i}_{t+1}, \ldots, \boldsymbol{i}_{t+h-1}\}$, then we can extend y to a row vector y' such that $\mathrm{supp}(y'A[t+\ell]) = \{v\}$, and thus $v \in \mathfrak{F}(A[t+h-1])$. Furthermore, since $(\boldsymbol{i}_{t+1}, \ldots, \boldsymbol{i}_{t+h-1}) \in [n]^{h-1}$ is uniformly random, we conclude that

$$\mathbb{P}\left[I = \{v, \boldsymbol{i}_{t+1}, \ldots, \boldsymbol{i}_{t+h-1}\} \,|A[t]\right] = (h-1)!/n^{h-1} \ge n^{1-h}. \tag{3.8}$$

Now, because every $v \in \mathscr{V}_{t,h}$ satisfies $r_{v,t,h} \ge \delta h n^{h-1}/(2\ell)$, (3.8) implies that

$$\mathbb{P}\left[v \in \mathfrak{F}(A[t+h-1])|A[t]\right] \ge r_{v,t,h}/n^{h-1} \ge \delta h/(2\ell). \tag{3.9}$$

We also notice that $\mathscr{V}_{t,h} \cap \mathfrak{F}(A[t]) = \emptyset$ because no minimal h-relation contains a frozen variable. Therefore, combining (3.1), (3.7), and (3.9) and using linearity of expectation, we obtain

$$\Delta_t \ge \frac{1}{n} \sum_{v \in \mathscr{V}_{t,h}} \mathbb{P}\left[v \in \mathfrak{F}(A[t+h-1])|A[t]\right] \ge \frac{\delta h |\mathscr{V}_{t,h}|}{2\ell n} \ge \frac{\delta^2 h^2}{4\ell^2} \ge \frac{\delta^2}{\ell^2},$$

as desired. ∎

Combining Lemmas 3.3 and 3.4, we immediately obtain the following.

Corollary 3.5. *If $A[t]$ fails to be (δ,ℓ)-free then $\Delta_t \ge \delta^2/\ell^2$.*

We have all the ingredients in place to complete the proof of Proposition 2.4.

Proof of Proposition 2.4. We define $T = \{t \in [\mathscr{T}\,] : \mathbb{P}[A[t] \text{ fails to be } (\delta, \ell)-\text{free}] \geq \delta/2\}$ so that

$$\mathbb{P}[A[\boldsymbol{\theta}] \text{ is } (\delta, \ell) - \text{free}] > 1 - \delta/2 - \mathbb{P}[\boldsymbol{\theta} \in T]. \tag{3.10}$$

Hence, we are left to estimate $\mathbb{P}[\boldsymbol{\theta} \in T]$. Applying Corollary 3.5, we obtain for every $t \in T$,

$$\mathbb{E}[\Delta_t] \geq \frac{\delta^2}{\ell^2} \cdot \mathbb{P}[A[t] \text{ fails to be } (\delta, \ell) - \text{free}] \geq \frac{\delta^3}{2\ell^2}. \tag{3.11}$$

Moreover, averaging (3.11) on $t \in [\mathscr{T}\,]$ and applying Lemma 3.1, we obtain

$$\frac{\delta^3}{2\ell^2} \cdot \mathbb{P}[\boldsymbol{\theta} \in T] = \frac{\delta^3}{2\ell^2} \cdot \frac{|T|}{\mathscr{T}} \leq \frac{1}{\mathscr{T}} \sum_{t \in T} \mathbb{E}[\Delta_t] \leq \mathbb{E}[\Delta_{\boldsymbol{\theta}}] \leq \frac{\ell}{\mathscr{T}}.$$

Consequently, choosing

$$\mathscr{T} > 4\ell^3/\delta^4, \tag{3.12}$$

ensures $\mathbb{P}[\boldsymbol{\theta} \in T] \leq \delta/2$. Thus, the assertion follows from (3.10). ∎

Remark 3.6. The proof presented in this section actually renders a slightly stronger statement than Proposition 2.4. Specifically, let A be an $m \times N$-matrix and let $n \leq N$. Obtain $A[\theta, n]$ by pegging θ random variables from among the first n variables $x_1, \ldots, x_n$ of the linear system $Ax = 0$ to zero. Then with $\boldsymbol{\theta} = \boldsymbol{\theta}(\delta, \ell)$ chosen as in Proposition 2.4 we find that with probability at least $1 - \delta$, there are no more than δn^{ℓ} proper relations $I \subseteq [n]$. Thus, in order to rid a subset of the columns of short linear relations, it suffices to peg $\boldsymbol{\theta}$ random variables from that subset to zero. The proof of this stronger statement proceeds as above, except that we confine ourselves to minimal relations among the first n columns.

3.2 | Proof of Lemma 2.5

We are going to derive Lemma 2.5 from the following simpler, deterministic and nonasymptotic statement.

Lemma 3.7. *Let $m, n, m', n' \geq 1$ be integers. Let A be an $m \times n$ matrix, let B be an $m' \times n$ matrix and let C be an $m' \times n'$ matrix. Let $I \subseteq [n]$ be the set of all indices of nonzero columns of B. Unless I is a relation of A we have*

$$\operatorname{nul} A - \operatorname{nul} \begin{pmatrix} A & 0 \\ B & C \end{pmatrix} = \operatorname{rk}(B\ C) - n'.$$

Proof. Suppose that I is not a relation of A. We begin by showing that

$$\operatorname{nul} A - \operatorname{nul} \begin{pmatrix} A \\ B \end{pmatrix} = \operatorname{rk}(B). \tag{3.13}$$

Writing $B_1, \dots, B_{m'}$ for the rows of B and $r = \mathrm{rk}(B)$ for the rank and applying a row permutation if necessary, we may assume that $B_1, \dots, B_r$ are linearly independent. Hence, to establish (3.13) it suffices to prove that for all $0 \le \ell < r$,

$$\mathrm{rk}\begin{pmatrix} A \\ B_1 \\ \vdots \\ B_\ell \end{pmatrix} < \mathrm{rk}\begin{pmatrix} A \\ B_1 \\ \vdots \\ B_{\ell+1} \end{pmatrix}. \tag{3.14}$$

In other words, we need to show that $B_{\ell+1}$ does not belong to the space spanned by $B_1, \dots, B_\ell$ and the rows $A_1, \dots, A_m$ of A. Indeed, assume that $B_{\ell+1} = \sum_{i=1}^{\ell} x_i B_i + \sum_{i=1}^{m} y_i A_i$. Then $0 \neq B_{\ell+1} - \sum_{i=1}^{\ell} x_i B_i = \sum_{i=1}^{m} y_i A_i$ and thus $\emptyset \neq \mathrm{supp} \sum_{i=m}^{\ell} y_i A_i \subseteq I$, in contradiction to the assumption that I is no relation of A. Hence, we obtain (3.14) and thus (3.13). Finally, to complete the proof of (2.4) we apply (3.13) to the matrices $(A\ 0)$ and $(B\ C)$, obtaining

$$\mathrm{nul}(A) + n' - \mathrm{nul}\begin{pmatrix} A & 0 \\ B & C \end{pmatrix} = \mathrm{nul}(A\ 0) - \mathrm{nul}\begin{pmatrix} A & 0 \\ B & C \end{pmatrix} = \mathrm{rk}(B\ C),$$

as desired. ∎

Proof of Lemma 2.5. Recall that A has size $m \times n$. By Definition 2.1 a coordinate i is frozen iff the vector $e^{(i)} \in \mathbb{F}^{1\times n}$ whose ith entry equals one and whose other entries equal zero can be written as a linear combination of the rows of A. For every $i \in \mathfrak{F}(A)$ we can therefore apply elementary row operations (like in Gaussian elimination) to zero out the entire i-column of B. Since elementary row operations do not alter the nullity of a matrix, we therefore obtain the identity

$$\mathrm{nul}\begin{pmatrix} A & 0 \\ B & C \end{pmatrix} = \mathrm{nul}\begin{pmatrix} A & 0 \\ B_* & C \end{pmatrix}.$$

The assertion thus follows from Lemma 3.7. ∎

4 | CONCENTRATION

The principal aim of this section is to prove Proposition 2.6, that is, to argue that the rank of the actual matrix $\boldsymbol{A}_n$ that does not have any cavities and whose Tanner graph is simple is close to the expected rank of $\boldsymbol{A}_{\varepsilon,n}$ a.a.s. In other words, we need to show that the rank of a random matrix is sufficiently concentrated that conditioning on

$$\mathscr{D} = \left\{ \sum_{i=1}^{n} \boldsymbol{d}_i = \sum_{i=1}^{\boldsymbol{m}} \boldsymbol{k}_i \right\},$$

and on the event $\mathscr{S}$ that the Tanner graph is simple is inconsequential. The main tool will be the following local limit theorem for sums of independent random variables, which we use in Section 4.1 to calculate the probability of $\mathscr{D}$.

Theorem 4.1 ([34, p. 130]). *Suppose that $(\boldsymbol{X}_i)_{i\geq 1}$ is a sequence of i.i.d. variables that take values in $\mathbb{Z}$ such that the greatest common divisor of the support of $\boldsymbol{X}_1$ is one. Also assume that* $\mathrm{Var}[\boldsymbol{X}_1] = \sigma^2 \in (0,\infty)$. *Then*

$$\lim_{n\to\infty}\sup_{z\in\mathbb{Z}}\left|\sqrt{n}\mathbb{P}\left[\sum_{i=1}^{n}\boldsymbol{X}_i = z\right] - \frac{\exp(-(z-n\mathbb{E}[\boldsymbol{X}_1])^2/(2n\sigma^2))}{\sqrt{2\pi}\sigma}\right| = 0.$$

Subsequently, in Section 4.2 we calculate the probability of the event $\mathcal{S}$, proving Proposition 1.10 along the way. Finally, in Section 4.3 we complete the proof of Proposition 2.6.

4.1 | The event $\mathscr{D}$

Because $\mathbb{E}[\boldsymbol{d}^r] + \mathbb{E}[\boldsymbol{k}^r] < \infty$ for an $r > 2$, the event

$$\mathscr{M} = \left\{\max_{i\in[n]}\boldsymbol{d}_i + \max_{i\in[\boldsymbol{m}]}\boldsymbol{k}_i \leq \sqrt{n}/\log^9 n\right\} \qquad \text{satisfies} \qquad \mathbb{P}[\mathscr{M}] = 1 - o_n(1). \tag{4.1}$$

As an application of Theorem 4.1 we obtain the following estimate.

Lemma 4.2. *If* $\gcd(\boldsymbol{k})$ *divides n, then* $\mathbb{P}[\mathscr{D}] = \Theta_n(n^{-1/2})$ *and* $\mathbb{P}[\mathscr{D}\,|\mathscr{M}] = \Theta_n(n^{-1/2})$.

Proof. For $\mathbb{P}[\mathscr{D}\,|\mathscr{M}]$ there are several cases to consider. First, that $\mathrm{Var}(\boldsymbol{d}) = \mathrm{Var}(\boldsymbol{k}) = 0$, that is, $\boldsymbol{d}, \boldsymbol{k}$ are both atoms. Since $\boldsymbol{m}$ is a Poisson variable with mean dn/k we find $\mathbb{P}[\mathscr{D}\,|\mathscr{M}] = \mathbb{P}\left[\boldsymbol{m} = dn/k\right] = \Theta_n(n^{-1/2})$.

Second, suppose that $\mathrm{Var}(\boldsymbol{d}) > 0$ but $\mathrm{Var}(\boldsymbol{k}) = 0$. Then Theorem 4.1 and (4.1) show that

$$\mathbb{P}\left[\left|dn - \sum_{i=1}^{n}\boldsymbol{d}_i\right| \leq \sqrt{n}\wedge k \ \text{ divides } \ \sum_{i=1}^{n}\boldsymbol{d}_i\middle|\mathscr{M}\right] = \Omega_n(1). \tag{4.2}$$

Further, given $\left|dn - \sum_{i=1}^{n}\boldsymbol{d}_i\right| \leq \sqrt{n}$ and given k divides $\sum_{i=1}^{n}\boldsymbol{d}_i$, the event $k\boldsymbol{m} = \sum_{i=1}^{n}\boldsymbol{d}_i$ has probability $\Theta_n(n^{-1/2})$ by the local limit theorem for the Poisson distribution.

The case that $\mathrm{Var}(\boldsymbol{d}) = 0$ but $\mathrm{Var}(\boldsymbol{k}) > 0$ can be dealt with similarly. Indeed, pick a large enough number $L > 0$ and let $I = \{i \in [\boldsymbol{m}] : \boldsymbol{k}_i > L\}$, $\boldsymbol{m}' = |I|$, $\boldsymbol{m}'' = \boldsymbol{m} - |I|$, $S' = \sum_{i\in I}\boldsymbol{k}_i$ and $S'' = \sum_{i\in[\boldsymbol{m}]\setminus I}\boldsymbol{k}_i$. Then $\boldsymbol{m}', \boldsymbol{m}''$ are stochastically independent, as are S', S''. Moreover, since S' satisfies the central limit theorem we have

$$\mathbb{P}\left[|S' - \mathbb{E}[S'|\mathscr{M}]| \leq \sqrt{n}|\mathscr{M}\right] = \Omega_n(1). \tag{4.3}$$

Further, Theorem 4.1 applies to S'', which is distributed as $\sum_{i=1}^{\boldsymbol{m}}\boldsymbol{k}_i\mathbf{1}\{\boldsymbol{k}_i \leq L\}$. Hence, as n is divisible by $\gcd(\boldsymbol{k})$, for large enough L we have

$$\mathbb{P}\left[S' + S'' = dn \| S' - \mathbb{E}[S'|\mathscr{M}]| \leq \sqrt{n}, \mathscr{M}\right] = \Omega_n(n^{-1/2}). \tag{4.4}$$

Thus, (4.3) and (4.4) show that $\mathbb{P}[\mathscr{D}\,|\mathscr{M}] = \Omega_n(n^{-1/2})$. The upper bound $\mathbb{P}[\mathscr{D}\,|\mathscr{M}] = O_n(n^{-1/2})$ follows from the uniform upper bound from Theorem 4.1.

A similar argument applies in the final case $\mathrm{Var}(\boldsymbol{d}) > 0$, $\mathrm{Var}(\boldsymbol{k}) > 0$. Indeed, Theorem 4.1 and (4.1) yield

$$\mathbb{P}\left[\gcd(\boldsymbol{k}) \text{ divides } \sum_{i=1}^{n} \boldsymbol{d}_i \text{ and } \left|dn - \sum_{i=1}^{n} \boldsymbol{d}_i\right| \le \sqrt{n} \mid \mathscr{M}\right] = \Omega_n(1). \tag{4.5}$$

Moreover, (4.3) remains valid regardless the variance of $\boldsymbol{d}$. Hence, applying Theorem 4.1 to S'', we obtain

$$\begin{aligned}\mathbb{P}\Bigg[S' + S'' = \sum_{i=1}^{n} \boldsymbol{d}_i \;\Big|\; \gcd(\boldsymbol{k}) \text{ divides } \sum_{i=1}^{n} \boldsymbol{d}_i, \; \left|dn - \sum_{i=1}^{n} \boldsymbol{d}_i\right| \\ \le \sqrt{n}, \; |S' - \mathbb{E}[S' \mid \mathscr{M}]| \le \sqrt{n}, \mathscr{M}\Bigg] = \Omega_n(n^{-1/2}).\end{aligned} \tag{4.6}$$

Combining (4.5) and (4.6), we see that $\mathbb{P}[\mathscr{D} \mid \mathscr{M}] = \Omega_n(n^{-1/2})$. The matching upper bound $\mathbb{P}[\mathscr{D} \mid \mathscr{M}] = O_n(n^{-1/2})$ follows from the universal upper bound from Theorem 4.1 once more. The treatment of the unconditional $\mathbb{P}[\mathscr{D}]$ is similar but slightly simpler. ∎

4.2 | The event $\mathscr{S}$

The random matrix $\boldsymbol{A}_n$ for Theorem 1.1 is identical in distribution to the random matrix $\boldsymbol{A}_{0,n}$ with $\varepsilon = 0$ conditioned on the event $\mathscr{D}$ and on the event $\mathscr{S}$ that the Tanner graph $\boldsymbol{G}_{0,n}$ does not contain any multi-edges. Therefore, Proposition 1.10 is going to be a consequence of Lemma 4.2and the following statement.

Lemma 4.3. *We have* $\mathbb{P}\left[\boldsymbol{A}_{0,n} \in \mathscr{S} \mid \mathscr{D}\right] = \Omega_n(1)$.

We proceed to prove Lemma 4.3. Recall the event $\mathscr{M}$ from (4.1). The proof of Lemma 4.3 is essentially based on the routine approach of showing by way of a moment calculation that the number of multi-edges of $\boldsymbol{G}_{0,n}$ is asymptotically Poisson with a finite mean. This argument has been carried out illustratively for the case of random regular graphs in [, chapter 9]. But since here we work with very general degree distributions, technical complications arise. For instance, as a first step we need to estimate the empirical variance of the degree sequences.

Claim 4.4. On the event $\mathscr{D} \cap \mathscr{M}$ we have $\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{d}_i^2 \to \mathbb{E}[\boldsymbol{d}^2]$, $\frac{1}{n}\sum_{i=1}^{\boldsymbol{m}} \boldsymbol{k}_i^2 \to d\mathbb{E}[\boldsymbol{k}^2]/k$ in probability.

Proof. We will only prove the statement about the $\boldsymbol{k}_i$; the same (actually slightly simplified) argument applies to the $\boldsymbol{d}_i$. Thanks to Bennett's tail bound for the Poisson distribution we may condition on $\{\boldsymbol{m} = m\}$ for some integer m with $|m - dn/k| \le \sqrt{n}\ln n$. Fix a small $\delta > 0$ and a large enough $L = L(\delta) > 0$. Given $\boldsymbol{m} = m$ the variables $Q_j = \sum_{i \in [\boldsymbol{m}]} \mathbf{1}\{\boldsymbol{k}_i = j\}$ have a binomial distribution. Therefore, the Chernoff bound yields

$$\mathbb{P}\left[\left|Q_j - dn\mathbb{P}[\boldsymbol{k} = j]/k\right| \le \sqrt{n}\ln n \mid \boldsymbol{m} = m\right] = 1 - o_n(1/n) \quad \text{for any } j \le L.$$

Hence, (4.1) and Lemma 4.2 yield

$$\mathbb{P}\left[\forall j \le L : |Q_j - dn\mathbb{P}[\boldsymbol{k} = j]/k| \le \sqrt{n}\ln n \mid \mathscr{D} \cap \mathscr{M}, \; \boldsymbol{m} = m\right] = 1 - o_n(1). \tag{4.7}$$

Further, let

$$R_h = \sum_{j\geq 1} \mathbf{1}\{(1+\delta)^{h-1}L < j \leq (1+\delta)^h L \wedge \sqrt{n}/\ln^9 n\} Q_j,$$
$$\overline{R}_h = m\sum_{j\geq 1} \mathbf{1}\{(1+\delta)^{h-1}L < j \leq (1+\delta)^h L \wedge \sqrt{n}/\ln^9 n\} \mathbb{P}[\boldsymbol{k} = j],$$

and let $\mathscr{H}$ be the set of all integers $h \geq 1$ with $(1+\delta)^{h-1}L \leq \sqrt{n}/\ln^9 n$. Then the Chernoff bound implies that

$$\mathbb{P}\left[\forall h \in \mathscr{H} : \left|R_h - \overline{R}_h\right| > \delta\overline{R}_h + \ln^2 n | \mathscr{D} \cap \mathscr{M}, \boldsymbol{m} = m\right] = o_n(n^{-1}). \tag{4.8}$$

Finally, if $|Q_j - dn\mathbb{P}[\boldsymbol{k} = j]/k| \leq \sqrt{n}\ln n$ for all $j \leq L$ and $\left|R_h - \overline{R}_h\right| \leq \delta\overline{R}_h + \ln^2 n$ for all $h \in \mathscr{H}$, then

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{m} \boldsymbol{k}_i^2 &\leq o_n(1) + \frac{d}{k}\mathbb{E}\left[\boldsymbol{k}^2\mathbf{1}\{\boldsymbol{k} \leq L\}\right] + \frac{d}{kn}\sum_{h\in\mathscr{H}}(1+\delta)^{2h}L^2 R_h \\ &= o_n(1) + \frac{d}{k}\mathbb{E}\left[\boldsymbol{k}^2\mathbf{1}\{\boldsymbol{k} \leq L\}\right] + \frac{d}{kn}\sum_{h\in\mathscr{H}}(1+\delta)^{2h+1}L^2\overline{R}_h \leq \frac{(1+\delta)d}{k}\mathbb{E}[\boldsymbol{k}^2] + o_n(1), \\ \text{and analogously } \frac{1}{n}\sum_{i=1}^{m} \boldsymbol{k}_i^2 &\geq \frac{(1-\delta)d}{k}\mathbb{E}[\boldsymbol{k}^2] + o_n(1).\end{aligned}$$

Since this holds true for any fixed $\delta > 0$, the assertion follows from (4.7) and (4.8). ∎

Claim 4.5. Let Y be the number of multi-edges of the Tanner graph $\boldsymbol{G}_{0,n}$ and let $\ell \geq 1$. There is $\lambda > 0$ such that on

$$\begin{aligned}\mathscr{M} \cap \mathscr{D} \cap \left\{\sum_{i=1}^{n} \boldsymbol{d}_i = dn + o_n(n),\ \sum_{i=1}^{n} \boldsymbol{d}_i^2 = n\mathbb{E}[\boldsymbol{d}^2] + o_n(n)\right\} \\ \cap \left\{\sum_{i=1}^{\boldsymbol{m}} \boldsymbol{k}_i^2 = dn\mathbb{E}[\boldsymbol{k}^2]/k + o_n(n)\right\} \cap \{\boldsymbol{m} = dn/k + o_n(n)\},\end{aligned}$$

we have

$$\mathbb{E}\left[\prod_{i=1}^{\ell} Y - i + 1 \;\middle|\; (\boldsymbol{d}_i)_{i\in[n]}, (\boldsymbol{k}_i)_{i\in\boldsymbol{m}}\right] = \lambda^\ell + o_n(1).$$

Proof. To estimate the ℓ-th factorial moments of Y for $\ell \geq 1$, we split the random variable into a sum of indicator variables. Specifically, let U_ℓ be the set of all families $(u_i, v_i, w_i)_{i\in\ell}$ with $u_i \in [\boldsymbol{m}]$, $v_i \in [n]$ and $2 \leq w_i \leq \boldsymbol{k}_{u_i} \wedge \boldsymbol{d}_{v_i} \leq \sqrt{n}/\log^9 n$ such that the pairs $(u_1, v_1), \ldots, (u_\ell, v_\ell)$ are pairwise distinct. Moreover, let $Y[(u_i, v_i, w_i)_{i\in[\ell]}]$ be the number of ordered ℓ-tuples of multi-edges comprising precisely w_i edges between check a_{u_i} and variable x_{v_i} for each i. Then

$$\prod_{h=1}^{\ell} Y - h + 1 = \sum_{(u_i,v_i,w_i)_{i\in[\ell]}\in U_\ell} Y[(u_i, v_i, w_i)_{i\in[\ell]}].$$

Moreover, letting $w = \sum_i w_i$, we claim that

$$\mathbb{E}[Y[(u_i, v_i, w_i)_{i\in[h]}]|(\boldsymbol{d}_i)_{i\in[n]}, (\boldsymbol{k}_i)_{i\in\boldsymbol{m}}] \sim \frac{1}{(dn)^w}\prod_{i=1}^{\ell}\binom{\boldsymbol{k}_{u_i}}{w_i}\binom{\boldsymbol{d}_{v_i}}{w_i}w_i!. \tag{4.9}$$

Indeed, the factors $\binom{\boldsymbol{d}_{v_i}}{w_i}\binom{\boldsymbol{k}_{u_i}}{w_i}w_i!$ count the number of possible matchings between w_i clones of the variable node x_{v_i}, whose degree equals $\boldsymbol{d}_{v_i}$, and of the check node a_{u_i} of degree $\boldsymbol{k}_{u_i}$. Further, since ℓ is bounded, the probability that all these matchings are realized in $\boldsymbol{G}_{0,n}$ is asymptotically equal to $(dn)^{-w}$.

Now, for a sequence $\boldsymbol{w} = (w_1, \ldots, w_\ell)$ let $Y_{\boldsymbol{w}} = \sum_{(u_i,v_i,w_i)_{i\in[\ell]}\in U_\ell} Y[(u_i, v_i, w_i)_{i\in[\ell]}]$. Then (4.9) yields

$$\begin{aligned}\mathbb{E}[Y_{\boldsymbol{w}}|(\boldsymbol{d}_i)_{i\in[n]}, (\boldsymbol{k}_i)_{i\in\boldsymbol{m}}] &\le O_n(n^{-w})\prod_{i=1}^{\ell}\left(\sum_{j=1}^{n}\boldsymbol{d}_j^{w_i}\right)\left(\sum_{j=1}^{\boldsymbol{m}}\boldsymbol{k}_j^{w_i}\right)\\ &\le O_n(n^{-w})\max_{j\in[n]}\boldsymbol{d}_j^{w-2\ell}\max_{j\in[\boldsymbol{m}]}\boldsymbol{k}_j^{w-2\ell}\left(\sum_{j=1}^{n}\boldsymbol{d}_j^2\right)^{\ell}\left(\sum_{j=1}^{\boldsymbol{m}}\boldsymbol{k}_j^2\right)^{\ell}\\ &\le O_n(n^{2\ell-w})\max_{j\in[n]}\boldsymbol{d}_j^{w-2\ell}\max_{j\in[\boldsymbol{m}]}\boldsymbol{k}_j^{w-2\ell} = O_n(\ln^{2\ell-w}n);\end{aligned}$$

the last bound follows from our conditioning on $\mathscr{M}$. As a consequence,

$$\sum_{\boldsymbol{w}:w>2\ell}\mathbb{E}[Y_{\boldsymbol{w}}|(\boldsymbol{d}_i)_{i\in[n]}, (\boldsymbol{k}_i)_{i\in\boldsymbol{m}}] = o_n(1). \tag{4.10}$$

Further, invoking (4.9), we obtain

$$\mathbb{E}[Y_{(2,\ldots,2)}|(\boldsymbol{d}_i)_{i\in[n]}, (\boldsymbol{k}_i)_{i\in\boldsymbol{m}}] \sim \lambda^\ell, \quad \text{where} \quad \lambda \sim \frac{\left(\sum_{i=1}^{n}\boldsymbol{d}_i(\boldsymbol{d}_i-1)\right)\left(\sum_{i=1}^{n}\boldsymbol{k}_i(\boldsymbol{k}_i-1)\right)}{2(dn)^2}. \tag{4.11}$$

Combining (4.10) and (4.11), we conclude that on $\mathscr{D}\cap\mathscr{M}\cap\{\boldsymbol{m} = dn/k + o_n(n)\}$,

$$\mathbb{E}[Y^\ell]|(\boldsymbol{d}_i)_{i\in[n]}, (\boldsymbol{k}_i)_{i\in\boldsymbol{m}}] \sim \lambda^\ell. \tag{4.12}$$

Finally, on $\left\{\sum_{i=1}^{n}\boldsymbol{d}_i = dn + o_n(n),\ \sum_{i=1}^{n}\boldsymbol{d}_i^2 = n\mathbb{E}[\boldsymbol{d}^2] + o_n(n)\right\}\cap\left\{\sum_{i=1}^{\boldsymbol{m}}\boldsymbol{k}_i^2 = dn\mathbb{E}[\boldsymbol{k}^2]/k + o_n(n)\right\}$ we have

$$\lambda \sim \lambda = \frac{(\mathbb{E}[\boldsymbol{d}^2]-d)(\mathbb{E}[\boldsymbol{k}^2]-k)}{2d^2}, \tag{4.13}$$

and the assertion follows from (4.12) and (4.13). ∎

Claim 4.6. We have $\mathbb{P}[\mathcal{S}|\mathscr{D}\cap\mathscr{M}] = \Omega_n(1)$.

Proof. Claims 4.4 and 4.5 show together with inclusion/exclusion (e.g., [13, theorem 1.22]) that a.a.s. on $\mathscr{M}\cap\mathscr{D}$,

$$\mathbb{P}\left[Y = 0|(\boldsymbol{d}_i)_{i\in[n]}, (\boldsymbol{k}_i)_{i\in\boldsymbol{m}}\right] = \exp(-\lambda) = \Omega_n(1).$$

Since $\mathcal{S} = \{Y = 0\}$, the assertion follows. ∎

Proof of Lemma 4.3. The assertion follows immediately from (4.1), Corollary 4.2 and Corollary 4.6. ∎

Proof of Proposition 1.10. The proposition is immediate from Lemmas 4.2 and 4.3. ∎

4.3 | Proof of Proposition 2.6

The random matrix $\boldsymbol{A}_n$ has n columns and $\boldsymbol{m} \sim \mathrm{Po}(dn/k)$ rows, with the column and row degrees drawn from the distributions $\boldsymbol{d}$ and $\boldsymbol{k}$. By comparison, $\boldsymbol{A}_{\varepsilon,n}$ has slightly fewer, namely $\boldsymbol{m}_{\varepsilon,n} \sim \mathrm{Po}((1-\varepsilon)dn/k)$ rows. One might therefore think that the proof of Proposition 2.6 is straightforward, as it appears that $\boldsymbol{A}_n$ is obtained from $\boldsymbol{A}_{\varepsilon,n}$ by simply adding another random $\mathrm{Po}(\varepsilon dn/k)$ rows. Since adding $O_{\varepsilon,n}(\varepsilon n)$ rows cannot reduce the nullity by more than $O_{\varepsilon,n}(\varepsilon n)$, the bound on $\mathbb{E}[\mathrm{nul}(\boldsymbol{A}_n)] - \mathbb{E}[\mathrm{nul}(\boldsymbol{A}_{\varepsilon,n})]$ appears to be immediate. But there is a catch. Namely, in constructing $\boldsymbol{A}_n$ we condition on the event $\mathscr{D} = \{\sum_{i=1}^n \boldsymbol{d}_i = \sum_{i=1}^{\boldsymbol{m}} \boldsymbol{k}_i\}$. Thus, $\boldsymbol{A}_{\varepsilon,n}$ does not have the same distribution as the top $\mathrm{Bin}(\boldsymbol{m}, 1-\varepsilon)$ rows of $\boldsymbol{A}_n$ since the conditioning might distort the degree distribution. We need to show that this distortion is insignificant. To this end, recall that $\boldsymbol{m}_{\varepsilon,n} \sim \mathrm{Po}((1-\varepsilon)dn/k)$.

Lemma 4.7. *A.a.s. we have*

$$\mathbb{P}\left[\left|\mathrm{nul}(\boldsymbol{A}_{\varepsilon,n}) - \mathbb{E}[\mathrm{nul}(\boldsymbol{A}_{\varepsilon,n})|\boldsymbol{m}_{\varepsilon,n}, (\boldsymbol{d}_i)_{i\geq 1}, (\boldsymbol{k}_i)_{i\geq 1}]\right| > \sqrt{n}\ln n|\boldsymbol{m}_{\varepsilon,n}, (\boldsymbol{d}_i)_{i\geq 1}, (\boldsymbol{k}_i)_{i\geq 1}\right] = o_n(1),$$

Proof. Lemma 1.11 shows that $\sum_{i=1}^n \boldsymbol{d}_i, \sum_{i=1}^{\boldsymbol{m}_{\varepsilon,n}} \boldsymbol{k}_i = O_{\varepsilon,n}(n)$ and $\sum_{i=1}^{\boldsymbol{m}_{\varepsilon,n}} \boldsymbol{k}_i \leq \sum_{i=1}^n \boldsymbol{d}_i$ with probability $1-o_n(n^{-1})$. Assuming that this is so, consider a filtration $(\mathfrak{A}_t)_{t\leq \sum_{i=1}^{\boldsymbol{m}_{\varepsilon,n}} \boldsymbol{k}_i}$ that reveals the random matching $\boldsymbol{\Gamma}_{\varepsilon,n}$ one edge at a time. Then

$$\left|\mathbb{E}[\mathrm{nul}(\boldsymbol{A}_{\varepsilon,n})|\mathfrak{A}_{t+1}, \boldsymbol{m}_{\varepsilon,n}, (\boldsymbol{d}_i)_{i\geq 1}, (\boldsymbol{k}_i)_{i\geq 1} - \mathbb{E}[\mathrm{nul}(\boldsymbol{A}_{\varepsilon,n})|\mathfrak{A}_t, \boldsymbol{m}_{\varepsilon,n}, (\boldsymbol{d}_i)_{i\geq 1}, (\boldsymbol{k}_i)_{i\geq 1}\right| \leq O_{\varepsilon,n}(1),$$

for all t. Therefore, the assertion follows from Azuma's inequality. ∎

Let $\boldsymbol{A}_{n,\mathscr{D}}$ be the conditional version of the random matrix $\boldsymbol{A}_{0,n}$ given $\mathscr{D}$. Thus, given $\sum_{i=1}^n \boldsymbol{d}_i = \sum_{i=1}^{\boldsymbol{m}} \boldsymbol{k}_i$, we construct a random Tanner multi-graph with variable degrees $\boldsymbol{d}_1, \ldots, \boldsymbol{d}_n$ and check degrees $\boldsymbol{k}_1, \ldots, \boldsymbol{k}_{\boldsymbol{m}}$. Hence, the difference between $\boldsymbol{A}_n$ and $\boldsymbol{A}_{n,\mathscr{D}}$ is merely that in the case of $\boldsymbol{A}_n$ we also condition on the event $\mathscr{S}$ that the Tanner graph is simple.

Lemma 4.8. *There exists a coupling of $\boldsymbol{A}_{n,\mathscr{D}}$ and $\boldsymbol{A}_{\varepsilon,n}$ such that with probability at least $1-\varepsilon$ the two matrices agree in all but $O_{\varepsilon,n}(\varepsilon n)$ rows.*

Proof. Let $\boldsymbol{G}_{n,\mathscr{D}}$ and $\boldsymbol{G}_{\varepsilon,n}$ denote the Tanner graphs corresponding to $\boldsymbol{A}_{n,\mathscr{D}}$ and $\boldsymbol{A}_{\varepsilon,n}$, respectively. It suffices to construct a coupling of $\boldsymbol{G}_{n,\mathscr{D}}$ and $\boldsymbol{G}_{\varepsilon,n}$ such that these graphs differ in edges incident with at most $O_{\varepsilon,n}(\varepsilon n)$ check nodes. To construct the coupling we first generate the following parameters for $\boldsymbol{A}_{\varepsilon,n}$. Parameter $\mathscr{T} = \mathscr{T}(\varepsilon)$ is given. Generate $\boldsymbol{\theta} \in [\mathscr{T}]$ uniformly at random. Then generate $\boldsymbol{m} \sim \mathrm{Po}(dn/k)$ and $\boldsymbol{m}_{\varepsilon,n} = \mathrm{Bin}(\boldsymbol{m}, 1-\varepsilon)$ and then check nodes $a_1, \ldots, a_{\boldsymbol{m}_{\varepsilon,n}}$. Each check node a_i is associated with an integer $\boldsymbol{k}_i$ which is an independent copy of $\boldsymbol{k}$. To distinguish $\boldsymbol{G}_{\varepsilon,n}$ from $\boldsymbol{G}_{n,\mathscr{D}}$, we colour these check nodes red. Add $\boldsymbol{\theta}$ check nodes $p_1, \ldots, p_{\boldsymbol{\theta}}$ to both $\boldsymbol{G}_{\varepsilon,n}$ and $\boldsymbol{G}_{n,\mathscr{D}}$.

Next generate n variable nodes where variable node x_i is associated with $\boldsymbol{d}_i$, which is an independent copy of $\boldsymbol{d}$. Further, let $\boldsymbol{r}_j = \sum_{h=1}^{\boldsymbol{m}} \mathbf{1}\{\boldsymbol{k}_h = j\}$ denote the prospective number of checks of $\boldsymbol{G}_{n,\mathscr{D}}$ of degree j. Applying Azuma's inequality and (4.1), we see that for any $\varepsilon > 0$ there exists $L = L(\varepsilon) > 0$ such that

$$\mathbb{P}\left[\sum_{j\geq L}\boldsymbol{r}_j > \varepsilon n|\mathscr{M}\right] + \mathbb{P}\left[\exists j \leq L : \boldsymbol{r}_j \leq \sum_{i=1}^{\boldsymbol{m}_{\varepsilon,n}} \mathbf{1}\{\boldsymbol{k}_i = j\}|\mathscr{M}\right] + \mathbb{P}\left[\boldsymbol{m} > \boldsymbol{m}_{\varepsilon,n} + 2\varepsilon dn/k\right] \leq 1/n.$$

Hence, Lemma 4.2 implies that

$$\begin{aligned}\mathbb{P}\left[\sum_{j\geq L}\boldsymbol{r}_j > \varepsilon n|\mathscr{D}\cap\mathscr{M}\right] + \mathbb{P}\left[\exists j \leq L : \boldsymbol{r}_j \leq \sum_{i=1}^{\boldsymbol{m}_{\varepsilon,n}} \mathbf{1}\{\boldsymbol{k}_i = j\} \text{ for all } j \leq L \,|\mathscr{D}\cap\mathscr{M}\right] \\ + \mathbb{P}\left[\boldsymbol{m} > \boldsymbol{m}_{\varepsilon,n} + 2\varepsilon dn/k|\mathscr{D}\cap\mathscr{M}\right] \leq 1/n = o_n(1).\end{aligned} \tag{4.14}$$

Now condition on the event

$$\mathscr{R} = \mathscr{D}\cap\mathscr{M}\cap\left\{\sum_{j\geq L}\boldsymbol{n}_j \leq \varepsilon n\right\}\cap\left\{\forall j \leq L : \boldsymbol{n}_j > \sum_{i=1}^{\boldsymbol{m}_{\varepsilon,n}} \mathbf{1}\{\boldsymbol{k}_i = j\}\right\}\cap\left\{\boldsymbol{m} \leq \boldsymbol{m}_{\varepsilon,n} + 2\varepsilon dn/k\right\}.$$

Uncolour all (red) check nodes a_i with $\boldsymbol{k}_i \leq L$. Moreover, for each $j \leq L$, generate $\boldsymbol{r}_j - \sum_{i=1}^{\boldsymbol{m}_{\varepsilon,n}} \mathbf{1}\{\boldsymbol{k}_i = j\}$ additional check nodes of degree j and colour them blue. Finally, for each $j > L$, generate $\boldsymbol{r}_j$ blue check nodes of degree j.

Now $\boldsymbol{G}_{\varepsilon,n}$ is generated by taking a random maximal matching from the clones of all *uncoloured* and *red* check nodes $\{a_i\}\times[\boldsymbol{k}_i]$ (excluding check nodes $p_1, \ldots, p_{\boldsymbol{\theta}}$) to the set of variable clones

$$\bigcup_{j=1}^{n}\{x_j\}\times[\boldsymbol{d}_j],$$

and then adding an edge between p_i and x_i for $1 \leq i \leq \boldsymbol{\theta}$. The Tanner graph $\boldsymbol{G}_{n,\mathscr{D}}$ is generated by removing all matching edges from the clones of the *red* check nodes, and removing edges between p_i and a_i for $1 \leq i \leq \boldsymbol{\theta}$, and then matching all clones of the *blue* check nodes to the remaining clones of the variable nodes. Finally, (4.14) ensures that with probability at least $1-\varepsilon$, the two Tanner graphs differ in no more than $O_{\varepsilon,n}(\varepsilon n)$ check nodes. ∎

Proof of Proposition 2.6. Assume that (2.8) is satisfied for $C > 0$ and fix $C' > C$ and a small enough $\delta > 0$. Then we find a small $0 < \varepsilon = \varepsilon(\delta) < \delta$ such that

$$\limsup_{n\to\infty}\mathbb{E}[\mathrm{nul}(\boldsymbol{A}_{\varepsilon,n})]/n \leq C + \delta.$$

Hence, combining Lemmas 4.7 and 4.8 and taking into account that changing a single row can alter the nullity by at most one, we conclude that

$$\mathbb{P}\left[\mathrm{nul}(\boldsymbol{A}_{n,\mathscr{D}})/n \leq C + O_{\varepsilon,n}(\varepsilon)\right] > 1 - \varepsilon + o_n(1). \tag{4.15}$$

Finally, combining (4.15) and Lemma 4.3, we conclude that

$$\mathbb{P}\left[\mathrm{nul}(\boldsymbol{A}_{n,\mathscr{D}})/n \leq C + O_{\varepsilon,n}(\varepsilon)|\mathscr{S}\right] > 1 - \delta + o_n(1), \tag{4.16}$$

provided that $\varepsilon = \varepsilon(\delta)$ is small enough. Since $\boldsymbol{A}_{n,\mathscr{D}}$ given $\mathscr{S}$ is identical to $\boldsymbol{A}_n$, the desired upper bound on the nullity of $\boldsymbol{A}_n$ follows from (4.16). The same argument renders the lower bound. ∎

5 | THE AIZENMAN–SIMS–STARR SCHEME: PROOF OF PROPOSITION 2.7

In this section we prove Proposition 2.7. As set out in Section 2.2, we are going to bound the difference of the nullities of $\boldsymbol{A}_{\varepsilon,n+1}$ and $\boldsymbol{A}_{\varepsilon,n}$ via Proposition 2.4 and Lemma 2.5. This requires a coupling of the random variables $\mathrm{nul}(\boldsymbol{A}_{\varepsilon,n+1})$ and $\mathrm{nul}(\boldsymbol{A}_{\varepsilon,n})$.

5.1 | The coupling

We begin by introducing a more fine-grained description of the random matrices $\boldsymbol{A}_{\varepsilon,n}$ and $\boldsymbol{A}_{\varepsilon,n+1}$ to facilitate the construction of the coupling. To this end, let $\boldsymbol{M} = (\boldsymbol{M}_j)_{j\geq 1}$ and $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_j)_{j\geq 1}$ be sequences of Poisson variables with means

$$\mathbb{E}[\boldsymbol{M}_j] = (1-\varepsilon)\mathbb{P}\left[\boldsymbol{k} = j\right] dn/k, \quad \mathbb{E}[\boldsymbol{\Delta}_j] = (1-\varepsilon)\mathbb{P}\left[\boldsymbol{k} = j\right] d/k. \tag{5.1}$$

All of these random variables are mutually independent and independent of $\boldsymbol{\theta}$ and the $(\boldsymbol{d}_i)_{i\geq 1}$. Further, let

$$\boldsymbol{M}_j^+ = \boldsymbol{M}_j + \boldsymbol{\Delta}_j, \quad \boldsymbol{m}_{\varepsilon,n} = \sum_{j\geq 1} \boldsymbol{M}_j, \quad \boldsymbol{m}_{\varepsilon,n}^+ = \sum_{j\geq 1} \boldsymbol{M}_j^+. \tag{5.2}$$

Since $\sum_{j\geq 1} \boldsymbol{M}_j \sim \mathrm{Po}((1-\varepsilon)dn/k)$, (5.2) is consistent with the earlier convention that $\boldsymbol{m}_{\varepsilon,n} \sim \mathrm{Po}((1-\varepsilon)dn/k)$.

The random vectors $(\boldsymbol{d}_1, \ldots \boldsymbol{d}_n), \boldsymbol{M}$ naturally define a random Tanner (multi-)graph $\boldsymbol{G}_{n,\boldsymbol{M}}$ with variable nodes $x_1, \ldots, x_n$ and check nodes $p_1, \ldots, p_{\boldsymbol{\theta}}$ and $a_{i,j}$, $i \geq 1, j \in [\boldsymbol{M}_i]$. Its edges are induced by a random maximal matching $\boldsymbol{\Gamma}_{n,\boldsymbol{M}}$ of the complete bipartite graph with vertex classes

$$\bigcup_{h=1}^{n} \{x_h\} \times [\boldsymbol{d}_h] \quad \text{and} \quad \bigcup_{i\geq 1} \bigcup_{j=1}^{\boldsymbol{M}_i} \{a_{i,j}\} \times [i].$$

Each matching edge $(x_h, s, a_{i,j}, t) \in \boldsymbol{\Gamma}_{n,\boldsymbol{M}}$ induces an edge between x_h and $a_{i,j}$ in the Tanner graph. In addition, there is an edge between p_i and x_i for every $i \in [\boldsymbol{\theta}]$.

To define the random matrix $\boldsymbol{A}_{n,\boldsymbol{M}}$ to go with $\boldsymbol{G}_{n,\boldsymbol{M}}$, let $\chi : [0,1]^2 \to \mathbb{F}^*$ be a measurable map and let $(\boldsymbol{\zeta}_{i,j}, \boldsymbol{\xi}_i)_{i,j\geq 1}$ be uniformly distributed on $[0,1]$, mutually independent and independent of all other randomness. [7] With the rows of $\boldsymbol{A}_{n,\boldsymbol{M}}$ indexed by the check nodes of $\boldsymbol{G}_{n,\boldsymbol{M}}$ and the columns indexed by the variable nodes, we define the matrix entries by letting

$$(\boldsymbol{A}_{n,\boldsymbol{M}})_{p_i,\boldsymbol{x}_h} = \mathbf{1}\{i = j\} \qquad (i \in [\boldsymbol{\theta}], h \in [n]),$$

$$(\boldsymbol{A}_{n,\boldsymbol{M}})_{a_{i,j},\boldsymbol{x}_h} = \chi_{\boldsymbol{\zeta}_{i,j},\boldsymbol{\xi}_h} \sum_{s=1}^{i} \sum_{t=1}^{\boldsymbol{d}_h} \mathbf{1}\left\{\{(x_h,t),(a_{i,j},s)\} \in \boldsymbol{\Gamma}_{n,\boldsymbol{M}}\right\} \qquad (i \geq 1, j \in [\boldsymbol{M}_i], h \in [n]).$$

The Tanner graph $\boldsymbol{G}_{n+1,\boldsymbol{M}^+}$ and its associated random matrix $\boldsymbol{A}_{n+1,\boldsymbol{M}^+}$ are defined analogously.

[7] Unfortunately at this point there does not seem to be an ideal notation for the matrix and its entries. Because the random vector $\boldsymbol{M}$ depends on n and to preserve the analogy with common random graphs notation, we denote the random Tanner graph by $\boldsymbol{G}_{n,\boldsymbol{M}}$ and its associated random matrix by $\boldsymbol{A}_{n,\boldsymbol{M}}$. At the same time, in line with linear algebra conventions, when indexing matrix entries we let the first index refer to the row of the matrix and the second index to the column. Since the variable n nodes correspond to the columns, a degree of incoherence seems unavoidable.

Lemma 5.1. *For any $\theta > 0$ we have* $\mathbb{E}[\mathrm{nul}(\boldsymbol{A}_{\varepsilon,n})] = \mathbb{E}[\mathrm{nul}(\boldsymbol{A}_{n,\boldsymbol{M}})]$, $\mathbb{E}[\mathrm{nul}(\boldsymbol{A}_{\varepsilon,n+1})] = \mathbb{E}[\mathrm{nul}(\boldsymbol{A}_{n+1,\boldsymbol{M}^+})]$.

Proof. We defined $\boldsymbol{A}_{\varepsilon,n}$ as the $\boldsymbol{m}_{\varepsilon,n} \times n$-matrix with target column and row degrees drawn from $\boldsymbol{d}$ and $\boldsymbol{k}$ independently with a $\boldsymbol{\theta} \times \boldsymbol{\theta}$ identity matrix affixed at top. In effect, because $\boldsymbol{m}_{\varepsilon,n}$ is a Poisson variable, the number of rows of with target degree i is distributed as $\boldsymbol{M}_i$, and these numbers are mutually independent. Hence, nul $\boldsymbol{A}_{\varepsilon,n}$ and nul $\boldsymbol{A}_{n,\boldsymbol{M}}$ are identically distributed. The same argument applies to $\boldsymbol{A}_{\varepsilon,n+1}$. ∎

Up to this point we merely introduced a new description of $\boldsymbol{A}_{\varepsilon,n}$ and $\boldsymbol{A}_{\varepsilon,n+1}$. To actually couple them we introduce a third random matrix whose nullity we can easily compare to $\mathrm{nul}(\boldsymbol{A}_{n,\boldsymbol{M}})$ and $\mathrm{nul}(\boldsymbol{A}_{n+1,\boldsymbol{M}^+})$. Specifically, let $\boldsymbol{\gamma}_i \geq 0$ be the number of checks $a_{i,j}$, $j \in [\boldsymbol{M}_i^+]$, adjacent to the last variable node x_{n+1} in $\boldsymbol{G}_{n+1,\boldsymbol{M}^+}$. Also let $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_i)_{i\geq 1}$ and set

$$\boldsymbol{M}_i^- = \max\{\boldsymbol{M}_i - \boldsymbol{\gamma}_i, 0\}. \tag{5.3}$$

In (5.3) the max is necessary because potentially $\boldsymbol{\gamma}_i$ might exceed $\boldsymbol{M}_i$ as $\boldsymbol{\gamma}_i$ might include some of the "extra" $\boldsymbol{\Delta}_i$ checks included in $\boldsymbol{G}_{n+1,\boldsymbol{M}^+}$. Consider the random Tanner graph $\boldsymbol{G}' = \boldsymbol{G}_{n,\boldsymbol{M}^-}$ induced by a random maximal matching $\boldsymbol{\Gamma}'$ of the complete bipartite graph with vertex classes

$$\bigcup_{h=1}^{n} \{x_h\} \times [\boldsymbol{d}_h] \quad \text{and} \quad \bigcup_{i\geq 1} \bigcup_{j=1}^{\boldsymbol{M}_i^-} \{a_{i,j}\} \times [i].$$

For each variable x_i, $i = 1, \ldots, n$, let $\mathscr{C}$ be the set of clones from $\bigcup_{i\in[n]} \{x_i\} \times [\boldsymbol{d}_i]$ that $\boldsymbol{\Gamma}_{n,\boldsymbol{M}^-}$ leaves unmatched. We call the elements of $\mathscr{C}$ *cavities*.

Now, obtain the Tanner graph $\boldsymbol{G}''$ from $\boldsymbol{G}'$ by adding new check nodes

$$a''_{i,j} \text{ with target degree } i \text{for each } i \geq 1,\ j \in [\boldsymbol{M}_i - \boldsymbol{M}_i^-]. \tag{5.4}$$

The new checks are joined by a random maximal matching $\boldsymbol{\Gamma}''$ of the complete bipartite graph with vertex classes

$$\mathscr{C} \qquad \text{and} \qquad \bigcup_{i\geq 1} \bigcup_{j\in[\boldsymbol{M}_i - \boldsymbol{M}_i^-]} \{a''_{i,j}\} \times [i],$$

that is, for each matching edge we insert a corresponding variable-check edge.

Analogously, obtain $\boldsymbol{G}'''$ by adding one variable node x_{n+1} as well as check nodes $a'''_{i,j}, i \geq 1, j \in [\boldsymbol{\gamma}_i]$ and $b'''_{i,j}$, $i \geq 1, j \in [\boldsymbol{M}_i^+ - \boldsymbol{M}_i^- - \boldsymbol{\gamma}_i]$ to $\boldsymbol{G}'$. The new checks are connected to $\boldsymbol{G}'$ via a random maximal matching $\boldsymbol{\Gamma}'''$ of the complete bipartite graph with vertex classes

$$\mathscr{C} \quad \text{and} \quad \bigcup_{i\geq 1} \left(\bigcup_{j\in[\boldsymbol{\gamma}_i]} \{a'''_{i,j}\} \times [i-1] \cup \bigcup_{j\in[\boldsymbol{M}_i^+ - \boldsymbol{M}_i^- - \boldsymbol{\gamma}_i]} \{b'''_{i,j}\} \times [i] \right).$$

For each matching edge we insert the corresponding variable-check edge and in addition each of the check nodes $a'''_{i,j}$ gets connected to x_{n+1} by exactly one edge.

Finally, we introduce the random matrices $\boldsymbol{A}', \boldsymbol{A}'', \boldsymbol{A}'''$ whose nonzero entries represent the edges of $\boldsymbol{G}', \boldsymbol{G}'', \boldsymbol{G}'''$. Recalling that $(\boldsymbol{\zeta}_{i,j}, \boldsymbol{\xi}_i)_{i,j\geq 1}$ are uniform on $[0, 1]$ and independent of everything else,

we additionally introduce independent random variables $(\boldsymbol{\zeta}'_{i,j}, \boldsymbol{\zeta}''_{i,j})_{i,j\geq 1}$, also uniform on $[0,1]$. With the rows and columns indexed by check and variable nodes, respectively, we define

$$\begin{aligned}
\boldsymbol{A}'_{p_i,j} &= \boldsymbol{A}''_{p_i,j} = \boldsymbol{A}'''_{p_i,j} = \mathbf{1}\{i=j\} \qquad (i\in[\boldsymbol{\theta}], j\in[n]),\\
\boldsymbol{A}'_{a_{i,j},x_h} &= \boldsymbol{A}''_{a_{i,j},x_h} = \boldsymbol{A}'''_{a_{i,j},x_h} = \chi_{\boldsymbol{\zeta}_{i,j},\boldsymbol{\xi}_h}\sum_{s=1}^{i}\sum_{t=1}^{\boldsymbol{d}_h}\mathbf{1}\left\{\{(x_h,t),(a_{i,j},s)\}\in\boldsymbol{\Gamma}'\right\} \qquad (i\geq 1, j\in[\boldsymbol{M}_i^-], h\in[n]),\\
\boldsymbol{A}''_{a''_{i,j},x_h} &= \chi_{\boldsymbol{\zeta}'_{i,j},\boldsymbol{\xi}_h}\sum_{s=1}^{i}\sum_{t=1}^{\boldsymbol{d}_h}\mathbf{1}\left\{\{(x_h,t),(a''_{i,j},s)\}\in\boldsymbol{\Gamma}''\right\} \qquad (i\geq 1, j\in[\boldsymbol{M}_i-\boldsymbol{M}_i^-, h\in[n]),\\
\boldsymbol{A}'''_{a'''_{i,j},x_h} &= \chi_{\boldsymbol{\zeta}'_{i,j},\boldsymbol{\xi}_h}\sum_{s=1}^{i-1}\sum_{t=1}^{\boldsymbol{d}_h}\mathbf{1}\left\{\{(x_h,t),(a'''_{i,j},s)\}\in\boldsymbol{\Gamma}'''\right\} \qquad (i\geq 1, j\in[\boldsymbol{\gamma}_i, h\in[n]),\\
\boldsymbol{A}'''_{b'''_{i,j},x_h} &= \chi_{\boldsymbol{\zeta}''_{i,j},\boldsymbol{\xi}_h}\sum_{s=1}^{i}\sum_{t=1}^{\boldsymbol{d}_h}\mathbf{1}\left\{\{(x_h,t),(b'''_{i,j},s)\}\in\boldsymbol{\Gamma}'''\right\} \qquad (i\geq 1, j\in[\boldsymbol{M}_i^+-\boldsymbol{M}_i^--\boldsymbol{\gamma}_i], h\in[n]).
\end{aligned}$$

In line with the strategy outlined in Section 2, this construction ensures that $\boldsymbol{A}''$ and $\boldsymbol{A}'''$ are obtained from $\boldsymbol{A}'$ by adding a bounded expected number of rows and, in the case of $\boldsymbol{A}'''$, one column. The following lemma links $\boldsymbol{A}'', \boldsymbol{A}'''$ to the random matrices $\boldsymbol{A}_{n,\boldsymbol{M}}$, $\boldsymbol{A}_{n+1,\boldsymbol{M}^+}$ from the beginning of the section.

Lemma 5.2. *We have* $\mathbb{E}[\mathrm{nul}(\boldsymbol{A}'')] = \mathbb{E}[\mathrm{nul}(\boldsymbol{A}_{n,\boldsymbol{M}})]+o_n(1)$ *and* $\mathbb{E}[\mathrm{nul}(\boldsymbol{A}''')] = \mathbb{E}[\mathrm{nul}(\boldsymbol{A}_{n+1,\boldsymbol{M}^+})]+o_n(1)$.

The proof of Lemma 5.2, deferred to Section 5.5, is tedious but relatively straightforward.

As a next step we are going to calculate the differences $\mathrm{nul}(\boldsymbol{A}''')-\mathrm{nul}(\boldsymbol{A}')$ and $\mathrm{nul}(\boldsymbol{A}'')-\mathrm{nul}(\boldsymbol{A}')$. We obtain expressions of one parameter of $\boldsymbol{A}'$, namely the fraction of cavities "frozen" to zero. To be precise, a cavity $(x_i,h)\in\mathscr{C}$ is *frozen* if $x_i\in\mathfrak{F}(\boldsymbol{A}')$. Let $\mathscr{F}\subseteq\mathscr{C}$ be the set of all frozen cavities and define $\boldsymbol{\alpha} = |\mathscr{F}|/|\mathscr{C}|$; in the unlikely event that $\mathscr{C}=\emptyset$, we agree that $\boldsymbol{\alpha}=0$. In Sections 5.3 and 5.4 we are going to establish the following two estimates.

Lemma 5.3. *We have* $\mathbb{E}[\mathrm{nul}(\boldsymbol{A}''')-\mathrm{nul}(\boldsymbol{A}')] = \mathbb{E}[D(1-K'(\boldsymbol{\alpha})/k)+d(K'(\boldsymbol{\alpha})+K(\boldsymbol{\alpha})-1)/k]-d+o_\varepsilon(1)$.

Lemma 5.4. *We have* $\mathbb{E}[\mathrm{nul}(\boldsymbol{A}'')-\mathrm{nul}(\boldsymbol{A}')] = d\mathbb{E}[\boldsymbol{\alpha}K'(\boldsymbol{\alpha})]/k-d+o_\varepsilon(1)$.

We emphasize that the r.h.s. expressions in Lemmas 5.3 and 5.4 involve expectations on the random variable $\boldsymbol{\alpha}$. A key feature of the present argument is that we manage to avoid an analysis of $\boldsymbol{\alpha}$ altogether. This is because, as the following proof of Proposition 2.7 shows, we can just replace the difference of the expectations by the largest conceivable value.

Proof of Proposition 2.7. Combining Lemmas 5.1 and 5.2, we see that

$$\begin{aligned}
\mathbb{E}[\mathrm{nul}(\boldsymbol{A}_{\varepsilon,n+1})]-\mathbb{E}[\mathrm{nul}(\boldsymbol{A}_{\varepsilon,n})] &= \mathbb{E}[\mathrm{nul}(\boldsymbol{A}_{n+1,\boldsymbol{M}^+})]-\mathbb{E}[\mathrm{nul}(\boldsymbol{A}_{n,\boldsymbol{M}})]\\
&= \mathbb{E}[\mathrm{nul}(\boldsymbol{A}''')]-\mathbb{E}[\mathrm{nul}(\boldsymbol{A}'')]+o_n(1)\\
&= (\mathbb{E}[\mathrm{nul}(\boldsymbol{A}''')]-\mathbb{E}[\mathrm{nul}(\boldsymbol{A}')])\\
&\quad-(\mathbb{E}[\mathrm{nul}(\boldsymbol{A}'')]-\mathbb{E}[\mathrm{nul}(\boldsymbol{A}')])+o_n(1). \qquad (5.5)
\end{aligned}$$

Further, combining (5.5) with Lemmas 5.3 and 5.4, we obtain

$$\begin{aligned}\mathbb{E}[\mathrm{nul}(\boldsymbol{A}_{\varepsilon,n+1})] - \mathbb{E}[\mathrm{nul}(\boldsymbol{A}_{\varepsilon,n})] &\le \mathbb{E}[D(1-K'(\boldsymbol{\alpha})/k) + d(K'(\boldsymbol{\alpha}) + K(\boldsymbol{\alpha}) - 1)/k - d\boldsymbol{\alpha}K'(\boldsymbol{\alpha})]/k + o_\varepsilon(1)\\ &= \mathbb{E}[\Phi(\boldsymbol{\alpha})] + o_\varepsilon(1) \le \max_{\alpha\in[0,1]} \Phi(\alpha) + o_\varepsilon(1). \end{aligned} \tag{5.6}$$

The proposition is an immediate consequence of (5.6). ∎

While proving Lemmas 5.3 and 5.4 in full detail requires a fair bit of work because we are dealing with very general degree distributions $\boldsymbol{d},\boldsymbol{k}$, it is not at all difficult to fathom where the right hand side expressions come from. They actually arise naturally from Lemma 2.5 and the scarcity of short proper relations supplied by Proposition 2.4. Indeed, we can write the matrices $\boldsymbol{A}'', \boldsymbol{A}'''$ in the form

$$\boldsymbol{A}'' = \begin{pmatrix}\boldsymbol{A}'\\ \boldsymbol{B}\end{pmatrix}, \quad \boldsymbol{A}''' = \begin{pmatrix}\boldsymbol{A}' & 0\\ \boldsymbol{B}' & \boldsymbol{C}'\end{pmatrix}, \tag{5.7}$$

with $\boldsymbol{B}, (\boldsymbol{B}'\ \boldsymbol{C}')$ representing the new rows and, in the case of $\boldsymbol{A}'''$, the additional column. To calculate $\mathbb{E}[\mathrm{nul}(\boldsymbol{A}''[\boldsymbol{\theta}]) - \mathrm{nul}(\boldsymbol{A}'[\boldsymbol{\theta}])]$ we basically need to assess the impact of adding a few more checks $a''_{i,j}$ to the Tanner graph $\boldsymbol{G}'$ of $\boldsymbol{A}'$. The new checks connect to randomly chosen cavities of $\boldsymbol{A}'$. Let $k_1, \ldots, k_L$ denote the degrees of the new checks. Since the distribution $\boldsymbol{k}$ of the check degrees has a finite second moment, the total degree $k_1 + \cdots + k_L$ is bounded a.a.s. The random matrix $\boldsymbol{B}$ therefore encodes the non-zero entries corresponding to the edges that connect the $a'_{i,j}$ with the cavities of $\boldsymbol{A}'$ where the new checks attach. Furthermore, the construction of $\boldsymbol{A}'$ ensures that a.a.s. the number of cavities is as large as $(1+o_n(1))\varepsilon dn$, and the $a'_{i,j}$ hatch on to randomly chosen cavities. Therefore, Proposition 2.4, applied with $\mathscr{T} = \mathscr{T}(\varepsilon)$ large enough, implies that the probability that the set I of nonzero columns of $\boldsymbol{B}$ forms a proper relation of $\boldsymbol{A}'$ is $o_\varepsilon(1)$. Consequently, Lemma 2.5 yields

$$\mathbb{E}[\mathrm{nul}(\boldsymbol{A}'') - \mathrm{nul}(\boldsymbol{A}')] = -\mathbb{E}[\mathrm{rk}(\boldsymbol{B}_*)] + o_n(1), \tag{5.8}$$

where $\boldsymbol{B}_*$ is obtained from $\boldsymbol{B}$ by zeroing out all columns indexed by $\mathfrak{F}(\boldsymbol{A}')$. Further, since the number of cavities of $\boldsymbol{A}'$ is as large as $\Omega_n(n)$ while $k_1 + \cdots + k_L = o_n(\sqrt{n})$ a.a.s., the matrix $\boldsymbol{B}$ has the following form a.a.s.: there are L rows containing $k_1, \ldots, k_L$ nonzero entries, respectively, and every column of $\boldsymbol{B}$ contains at most one nonzero entry. Consequently, once more because there are as many as $\Omega_n(n)$ cavities out of which an $\boldsymbol{\alpha}$ fraction are frozen to zero, $\boldsymbol{B}_*$ is close in total variation to the matrix obtained from $\boldsymbol{B}$ by zeroing out every column with probability $\boldsymbol{\alpha}$ independently. In effect, the probability that the i-th row of $\boldsymbol{B}_*$ gets zeroed out entirely is $\boldsymbol{\alpha}^{k_i} + o_n(1)$. Thus, a.a.s. we have

$$\mathbb{E}[\mathrm{rk}(\boldsymbol{B}_*)|\boldsymbol{\alpha}, k_1, \ldots, k_L] = \sum_{i=1}^{L} \left(1 - \boldsymbol{\alpha}^{k_i}\right) + o_{\varepsilon,n}(1). \tag{5.9}$$

Substituting (5.9) into (5.8) and the correct distribution of $k_1, \ldots, k_L$ supplied by the coupling into (5.9), we obtain the expression displayed in Lemma 5.4. To be explicit, the correct degrees $k_1, \ldots, k_L$ are provided by (5.4), that is, there are $\boldsymbol{M}_i - \boldsymbol{M}_i^-$ checks of degree i for every i. Hence, to obtain the expression in Lemma 5.3 we need to analyze the random variables $\boldsymbol{\gamma}_i$ from (5.3). This analysis will be conducted in Lemma 5.8 below, which shows that the $\boldsymbol{\gamma}_i$ are well approximated by the $\hat{\boldsymbol{\gamma}}_i$ from (5.15), which in turn come in terms of the reweighted check degree distribution from (1.10). A similar but slightly more complicated calculation explains the expression in Lemma 5.3. We proceed to prove Lemmas 5.2–5.4 formally. This requires a bit of groundwork.

5.2 | Groundwork

Let $P = P_{G'}$ be the distribution on the set $V_n = \{x_1, \ldots, x_n\}$ of variables induced by choosing a cavity uniformly at random, that is,

$$P(x_i) = |\mathscr{C} \cap (\{x_i\} \times [\boldsymbol{d}_i])|/|\mathscr{C}|;$$

in the (unlikely) event that $\mathscr{C} = \emptyset$, we use the convention $P(x_1) = 1$. Let $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots \in V_n$ be independent samples drawn from P. The following lemma shows that $|\mathscr{C}|$ is linear in n a.a.s.

Lemma 5.5. *A.a.s. we have* $|\mathscr{C}| \geq \varepsilon dn/2$.

Proof. The choice (5.1) of $\boldsymbol{M}$ ensures that $\mathbb{E}\sum_{j\geq 1} j\boldsymbol{M}_j = (1-\varepsilon)dn$. Moreover, because the $\boldsymbol{M}_j$ are mutually independent Poissons,

$$\mathrm{Var}\sum_{j\geq 1} j\boldsymbol{M}_j = \sum_{j\geq 1} j^2\mathrm{Var}(\boldsymbol{M}_j) = \sum_{j\geq 1} j^2\mathbb{E}[\boldsymbol{M}_j] = (1-\varepsilon)dn\mathbb{E}[\boldsymbol{k}^2]/k = O_{\varepsilon,n}(n).$$

Consequently, Chebyshev's inequality shows that

$$\mathbb{P}\left[\left|\sum_{j\geq 1} j\boldsymbol{M}_j - (1-\varepsilon)dn\right| \leq \sqrt{n}\log n\right] = 1 - o_n(1). \tag{5.10}$$

Similarly, we have $\mathbb{E}\sum_{i=1}^n \boldsymbol{d}_i = dn$ and $\mathrm{Var}\sum_{i=1}^n \boldsymbol{d}_i = \sum_{i=1}^n \mathrm{Var}(\boldsymbol{d}) = O_{\varepsilon,n}(n)$, whence

$$\mathbb{P}\left[\left|\sum_{i=1}^n \boldsymbol{d}_i - dn\right| \leq \sqrt{n}\log n\right] = 1 - o_n(1). \tag{5.11}$$

Since $|\mathscr{C}| \geq \sum_{i=1}^n \boldsymbol{d}_i - \sum_{j\geq 1} j\boldsymbol{M}_j$ by construction, the assertion follows from (5.10) and (5.11). ∎

Further, letting $\ell_* = \lceil \exp(1/\varepsilon^4)\rceil$ and $\delta_* = \exp(-1/\varepsilon^4)$, consider the event

$$\mathscr{E} = \left\{\mathbb{P}\left[\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{\ell_*} \text{ form a proper relation of } \boldsymbol{A}'|\boldsymbol{A}'\right] < \delta_*\right\}. \tag{5.12}$$

The following simple lemma is an application of Proposition 2.4.

Lemma 5.6. *For sufficiently large* $\mathscr{T} = \mathscr{T}(\varepsilon) > 0$ *we have* $\mathbb{P}\left[\boldsymbol{A}' \in \mathscr{E}\right] > \exp(-1/\varepsilon^4)$.

Proof. Lemma 5.5 provides that $|\mathscr{C}| \geq \varepsilon n/2$ a.a.s. Moreover, since $\mathbb{E}[\boldsymbol{d}] = O_{\varepsilon,n}(1)$ we find $L = L(\varepsilon) > 0$ such that the event $\mathscr{L} = \left\{\sum_{i=1}^n \boldsymbol{d}_i \mathbf{1}\{\boldsymbol{d}_i > L\} < \varepsilon\delta_*^2 n/(16\ell_*)\right\}$ has probability at least $1-\delta_*/8$. Thus, we may condition on the event $\mathscr{A} = \mathscr{L} \cap \{|\mathscr{C}| \geq \varepsilon n/2\}$.

Let $\hat{\boldsymbol{x}}_1, \ldots, \hat{\boldsymbol{x}}_{\ell_*}$ be a sequence of independently and uniformly chosen variables from $x_1, \ldots, x_n$. Consider a set $\mathscr{W} \subseteq \{x_1, \ldots, x_n\}^{\ell_*}$. How can we estimate the probability that $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{\ell_*}) \in \mathscr{W}$? Either one of the variables $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{\ell_*}$ has degree greater than L; on the event $\mathscr{A}$ this occurs with probability at most $\delta_*^2/16$. Or all of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{\ell_*}$ have degree at most L. Then the probability that $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{\ell_*}) \in \mathscr{W}$ is not much greater than the probability that $(\hat{\boldsymbol{x}}_1, \ldots, \hat{\boldsymbol{x}}_{\ell_*}) \in \mathscr{W}$. To be precise,

since $\hat{\boldsymbol{x}}_1, \ldots, \hat{\boldsymbol{x}}_{\ell_*}$ are chosen uniformly and there are at least $\varepsilon n/2$ cavities, the probabilities differ by no more than a factor of $(2L/\varepsilon)^{\ell_*}$. Hence, on the event $\mathscr{A}$ we have

$$\mathbb{P}\left[(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{\ell_*}) \in \mathscr{W} \mid \boldsymbol{A}'\right] \leq (2L/\varepsilon)^{\ell_*}\mathbb{P}\left[(\hat{\boldsymbol{x}}_1, \ldots, \hat{\boldsymbol{x}}_{\ell_*}) \in \mathscr{W} \mid \boldsymbol{A}'\right] + \delta_*^2/8. \tag{5.13}$$

Applying (5.13) to the set $\mathscr{W}$ of proper relations and invoking Proposition 2.4 completes the proof. ■

Further, consider the event

$$\mathscr{E}' = \left\{|\mathscr{C}| \geq \varepsilon dn/2 \wedge \max_{i \leq n} \boldsymbol{d}_i \leq n^{1/2}\right\}. \tag{5.14}$$

Lemma 5.7. *We have* $\mathbb{P}\left[\mathscr{E}'\right] = 1 - o_n(1)$.

Proof. This follows from the choice of the parameters in (5.1), Lemma 1.11 and Lemma 5.5. ■

To prove Lemmas 5.3 and 5.4 we need an explicit description of the vector $\boldsymbol{\gamma}$ that captures the degrees of the checks adjacent to the new variable node x_{n+1}. Since $\boldsymbol{\gamma}$ is defined in terms of the the "big" Tanner graph $\boldsymbol{G}_{n+1,\boldsymbol{M}^+}$, $\boldsymbol{\gamma}$ and the random variables are stochastically dependent. However, the next lemma shows that this dependence is very weak. Additionally, the lemma shows that the law of $\boldsymbol{\gamma}$ can be expressed easily in terms of the sequence $(\hat{\boldsymbol{k}}_i)_{i\geq 1}$ of independent copies of $\hat{\boldsymbol{k}}$ from (1.10). Indeed, let

$$\hat{\boldsymbol{\gamma}}_j = \sum_{i=1}^{\boldsymbol{d}_{n+2}} \mathbf{1}\{\hat{\boldsymbol{k}}_i = j\} \quad \text{and} \quad \hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\gamma}}_j)_{j\geq 1}. \tag{5.15}$$

Also let $\hat{\boldsymbol{\Delta}} = (\hat{\boldsymbol{\Delta}}_j)_{j\geq 1}$ be a family random variables, mutually independent and independent of everything else, with distributions

$$\hat{\boldsymbol{\Delta}}_j \sim \mathrm{Po}((1-\varepsilon)\mathbb{P}[\boldsymbol{k} = j]d/k). \tag{5.16}$$

Further, let Σ' be the σ-algebra generated by $\boldsymbol{G}'$, $\boldsymbol{A}'$, $\boldsymbol{M}^-$ and $(\boldsymbol{d}_i)_{i\in[n]}$. We write $\boldsymbol{\gamma}|\Sigma', \boldsymbol{\Delta}|\Sigma'$ for the conditional versions of $\boldsymbol{\gamma}, \boldsymbol{\Delta}$ given Σ'.

Lemma 5.8. *With probability* $1-\exp(-\Omega_{\varepsilon,n}(1/\varepsilon))$ *over the choice of* $\boldsymbol{G}', \boldsymbol{A}', \boldsymbol{M}^-$ *and* $(\boldsymbol{d}_i)_{i\in[n]}$ *we have*

$$d_{\mathrm{TV}}(\boldsymbol{\gamma}|\Sigma', \hat{\boldsymbol{\gamma}}) + d_{\mathrm{TV}}(\boldsymbol{\Delta}|\Sigma', \hat{\boldsymbol{\Delta}}) = O_{\varepsilon,n}(\varepsilon^{1/2}).$$

Proof. We begin by studying the unconditional distributions of $\boldsymbol{\gamma}$ and $\boldsymbol{\Delta}$.

Let $\zeta = (\sum_{i\geq 1} i\boldsymbol{M}_i^+)/(\sum_{i=1}^{n+1}\boldsymbol{d}_i)$. Proceeding as in the proof of Lemma 5.5, we conclude that $\mathbb{P}[1-2\varepsilon \leq \zeta \leq 1-\varepsilon/2] = 1 - o_n(1)$. Further, given $1-2\varepsilon \leq \zeta \leq 1-\varepsilon/2$ we can think of $\boldsymbol{G}_{n+1,\boldsymbol{M}^+}$ as being generated by the following experiment.

(i) Choose a set $\boldsymbol{C} \subseteq \bigcup_{h=1}^{n+1}\{x_h\} \times [\boldsymbol{d}_h]$ of size $(1-\zeta)\sum_{i=1}^{n+1}\boldsymbol{d}_i$ uniformly at random.
(ii) Create a random perfect matching $\boldsymbol{\Gamma}^\star$ of the complete bipartite graph with vertex classes

$$\left(\bigcup_{h=1}^{n+1}\{x_h\} \times [\boldsymbol{d}_h]\right) \setminus \boldsymbol{C} \quad \text{and} \quad \bigcup_{i\geq 1}\bigcup_{j=1}^{\boldsymbol{M}_i^+}\{a_{i,j}\} \times [i].$$

(iii) Obtain $\boldsymbol{G}^\star$ with variable nodes $x_1, \ldots, x_{n+1}$ and check nodes $a_{i,j}$, $i \geq 1, j \in [\boldsymbol{M}_i^+]$ by inserting an edge between x_h and $a_{i,j}$ for any edge of $\boldsymbol{\Gamma}^\star$ that links $\{x_h\} \times [\boldsymbol{d}_h]$ to $\{a_{i,j}\} \times [i]$.

In other words, in the first step we designate the set of $\mathscr{C} = \boldsymbol{C}$ of cavities and in the next two steps we connect the noncavities randomly.

By way of this alternative description we can easily get a grip on the degree of x_{n+1}. Indeed, given that $\boldsymbol{d}_{n+1} \leq \varepsilon^{-1/2}$, the probability that one of the clones $\{n+1\} \times [\boldsymbol{d}_{n+1}]$ ends up in $\boldsymbol{C}$ is $O_\varepsilon(\varepsilon^{1/2})$. Hence, the actual degree $\boldsymbol{d}^\star_{n+1}$ of x_{n+1} in $\boldsymbol{G}^\star$ satisfies

$$d_{\mathrm{TV}}\left(\boldsymbol{d}^\star_{n+1} | \{\boldsymbol{d}_{n+1} \leq \varepsilon^{-1/2}\}, \boldsymbol{d}\right) = O_{\varepsilon,n}(\varepsilon^{1/2}). \tag{5.17}$$

Regarding the degrees of the checks adjacent to x_{n+1}, by the principle of deferred decisions we can construct $\boldsymbol{\Gamma}^\star$ by matching one variable clone at a time, starting with the clones $\{x_{n+1}\} \times [\boldsymbol{d}_{n+1}]$. Clearly, in this process the probability that a specific clone of x_{n+1} links to a specific check is proportional to the degree of that check. Therefore, since $\mathbb{E}\sum_{i\geq 1} i\boldsymbol{M}_i^+ = O_{\varepsilon,n}(1)$, we find a fixed number L such that with probability $1 - O_{\varepsilon,n}(\varepsilon^{-1})$ all checks adjacent to x_{n+1} have degree at most L. Further, Chebyshev's inequality shows that $\boldsymbol{M}_i^+ = (1-\varepsilon)\mathbb{P}[\boldsymbol{k} = i]dn/k + o_n(n)$ for all $i \leq L$ and $\sum_{i\geq 1} i\boldsymbol{M}_i^+ = (1-\varepsilon)dn + o_n(n)$ a.a.s. In effect, if $\boldsymbol{d}_{n+1} \leq \varepsilon^{-1/2}$, the $\boldsymbol{d}_{n+1}$ choices of the checks are asymptotically independent, and the distribution of the individual check degrees that x_{n+1} joins is at total variation distance $o_n(1)$ of the distribution $\hat{\boldsymbol{k}}$. In summary, given $\boldsymbol{M}_i^+ = (1-\varepsilon)\mathbb{P}[\boldsymbol{k} = i]dn/k + o_n(n)$ for all $i \leq L$ and $\sum_{i\geq 1} i\boldsymbol{M}_i^+ = (1-\varepsilon)dn + o_n(n)$ we have

$$d_{\mathrm{TV}}(\boldsymbol{\gamma}, \hat{\boldsymbol{\gamma}}) = O_{\varepsilon,n}(\varepsilon^{1/2}). \tag{5.18}$$

Moreover, it is immediate from (5.1) that the unconditional $\boldsymbol{\Delta}$ is distributed as $\hat{\boldsymbol{\Delta}}$.

To complete the proof we are going to argue that $\boldsymbol{M}^-, \boldsymbol{d}_1, \ldots, \boldsymbol{d}_n$ and $\boldsymbol{\gamma}, \boldsymbol{\Delta}$ are asymptotically independent. Arguing along the lines of the previous paragraph, we find that for large $L = L(\varepsilon) > 0$ the event

$$\mathscr{K} = \left\{\sum_{i\geq 1} i(\boldsymbol{\Delta}_i + \boldsymbol{\gamma}_i) \leq L\right\},$$

occurs with probability $\mathbb{P}[\mathscr{K}] \geq 1 - \exp(-1/\varepsilon^2)$. Consequently, the event

$$\mathscr{L} = \{\mathbb{P}[\mathscr{K} | \boldsymbol{M}^-, \boldsymbol{d}_1, \ldots, \boldsymbol{d}_n] \geq 1 - \exp(-1/\varepsilon)\},$$

satisfies $\mathbb{P}[\mathscr{L}] \geq 1 - \exp(-1/\varepsilon)$. Moreover, since $\boldsymbol{M}$ comprises independent Poisson variables, the event

$$\begin{aligned}\mathscr{M} = &\left\{\forall i \leq L : |\boldsymbol{M}_i^- - \mathbb{E}[\boldsymbol{M}_i]| \leq \sqrt{n}\ln n\right\} \\ &\cap \left\{\sum_{i=1}^n \boldsymbol{d}_i = (1-\varepsilon)dn + o_n(n)\right\} \cap \left\{\sum_{i\geq 1} i\boldsymbol{M}_i^- = (1-\varepsilon)dn + o_n(n)\right\},\end{aligned}$$

satisfies $\mathbb{P}[\mathscr{M}] = 1 - o_n(1)$. In summary,

$$\mathbb{P}[\mathscr{K}] \geq \exp(-1/\varepsilon^2), \quad \mathbb{P}[\mathscr{L}] \geq 1 - \exp(-1/\varepsilon), \quad \mathbb{P}[\mathscr{M} | \mathscr{K}] = 1 - o_n(1). \tag{5.19}$$

Further, we claim that for any outcomes $(M^-, d_1, \ldots, d_n) \in \mathscr{L} \cap \mathscr{M}$ and $(\gamma, \Delta) \in \mathscr{K}$,

$$\mathbb{P}\left[\boldsymbol{\gamma} = \gamma, \boldsymbol{\Delta} = \Delta | \boldsymbol{M}^- = M^-, \forall i \in [n] : \boldsymbol{d}_i = d_i\right] \sim \mathbb{P}[\boldsymbol{\gamma} = \gamma]\,\mathbb{P}[\boldsymbol{\Delta} = \Delta]. \tag{5.20}$$

Indeed, on the event $\mathscr{M}$ we have $\boldsymbol{M}_i^- = \mathbb{E}[\boldsymbol{M}_i] + O_n\left(\sqrt{n}\ln n\right) = \Omega_n(n)$ for any $i \le L$ in the support of $\boldsymbol{k}$, the local limit theorem for the Poisson distribution yields

$$\begin{aligned}\mathbb{P}\left[\boldsymbol{M}^- = M^-, \forall i \le n : \boldsymbol{d}_i = d_i | \boldsymbol{\gamma} = \gamma, \boldsymbol{\Delta} = \Delta\right] &= \mathbb{P}\left[\boldsymbol{M} = M^- + \gamma, \forall i \le n : \boldsymbol{d}_i = d_i | \boldsymbol{\gamma} = \gamma, \boldsymbol{\Delta} = \Delta\right] \\ &= \frac{\mathbb{P}\left[\boldsymbol{\gamma} = \gamma, \boldsymbol{\Delta} = \Delta | \boldsymbol{M} = M^- + \gamma, \forall i \le n : \boldsymbol{d}_i = d_i\right]}{\mathbb{P}\left[\boldsymbol{\gamma} = \gamma, \boldsymbol{\Delta} = \Delta\right]} \cdot \mathbb{P}\left[\boldsymbol{M} = M^- + \gamma\right] \cdot \prod_{i=1}^{n} \mathbb{P}\left[\boldsymbol{d}_i = d_i\right] \\ &= (1 + o_n(1)) \frac{\mathbb{P}\left[\boldsymbol{\gamma} = \gamma | \boldsymbol{M} = M^- + \gamma, \forall i \le n : \boldsymbol{d}_i = d_i, \boldsymbol{\Delta} = \Delta\right]}{\mathbb{P}\left[\boldsymbol{\gamma} = \gamma\right]} \cdot \mathbb{P}\left[\boldsymbol{M} = M^-\right] \cdot \prod_{i=1}^{n} \mathbb{P}\left[\boldsymbol{d}_i = d_i\right]. \end{aligned} \tag{5.21}$$

Finally, given $\boldsymbol{M} = M^- + \gamma$ and $\boldsymbol{\Delta} = \Delta$ we have $\boldsymbol{M}_i^+ = (1-\varepsilon)\mathbb{P}[\boldsymbol{k} = i]dn/k + o_n(n)$ for all $i \le L$ and $\sum_{i \ge 1} i\boldsymbol{M}_i^+ = (1-\varepsilon)dn + o_n(n)$. Therefore, by the principle of deferred decisions, once we condition on a likely outcomes M^- of $\boldsymbol{M}^-, \boldsymbol{d}_1, \ldots, \boldsymbol{d}_n$ and of $\boldsymbol{\Delta}$, the conditional probability of obtaining $\boldsymbol{\gamma} = \gamma$ is close to the unconditional probability:

$$\mathbb{P}\left[\boldsymbol{\gamma} = \gamma | \boldsymbol{M} = M^- + \gamma, \forall i \le n : \boldsymbol{d}_i = d_i, \boldsymbol{\Delta} = \Delta\right] = (1 + o_n(1))\mathbb{P}\left[\boldsymbol{\gamma} = \gamma\right].$$

Hence, (5.20) follows from (5.19) and (5.21).

Finally, to complete the proof we combine (5.19) and (5.20) to conclude that with probability $1 - \exp(-\Omega_{\varepsilon,n}(1/\varepsilon))$,

$$\mathbb{P}\left[\boldsymbol{\gamma} = \gamma, \boldsymbol{\Delta} = \Delta | \Sigma'\right] = \mathbb{P}\left[\boldsymbol{\gamma} = \gamma, \boldsymbol{\Delta} = \Delta | \boldsymbol{M}^-, \boldsymbol{d}_1, \ldots, \boldsymbol{d}_n\right] = (1 + o_n(1))\mathbb{P}\left[\boldsymbol{\gamma} = \gamma\right]\mathbb{P}\left[\boldsymbol{\Delta} = \Delta\right]. \tag{5.22}$$

The assertion follows from (5.18) and (5.22). ∎

5.3 | Proof of Lemma 5.3

The proof comprises several steps, each relatively simple individually. Let

$$X = \sum_{i \ge 1} \boldsymbol{\Delta}_i, \quad Y = \sum_{i \ge 1} i\boldsymbol{\Delta}_i, \quad Y' = \sum_{i \ge 1} i\boldsymbol{\gamma}_i.$$

Then the total number of new nonzero entries upon going from $\boldsymbol{A}'$ to $\boldsymbol{A}'''$ is bounded by $Y + Y'$. Let

$$\mathscr{E}'' = \left\{X \vee Y \vee Y' \le 1/\varepsilon\right\}.$$

Claim 5.9. We have $\mathbb{P}\left[\mathscr{E}''\right] = 1 - O_{\varepsilon,n}(\varepsilon)$.

Proof. Since (5.1) yields $\mathbb{E}[X], \mathbb{E}[Y] = O_{\varepsilon,n}(1)$, Markov's inequality yields $\mathbb{P}[X > 1/\varepsilon] = O_{\varepsilon,n}(\varepsilon)$ and $\mathbb{P}[Y > 1/\varepsilon] = O_{\varepsilon,n}(\varepsilon)$. Further, we can bound the probability that a check of degree i is adjacent to x_{n+1} by $i\boldsymbol{d}_{n+1}/n$, because one of the i clones of the check has to be matched to one of the $\boldsymbol{d}_{n+1}$ clones of x_{n+1} and $\sum_{i=1}^{n} \boldsymbol{d}_i \ge n$. Hence,

$$\mathbb{E}\left[Y'\right] = \mathbb{E}\sum_{i \ge 1} i\boldsymbol{\gamma}_i \le \mathbb{E}\sum_{i \in [\boldsymbol{m}_{\varepsilon,n}^+]} \boldsymbol{k}_i^2 \boldsymbol{d}_{n+1}/n = O_{\varepsilon,n}(1).$$

Thus, the assertion follows from Markov's inequality. ∎

Going from $\boldsymbol{G}'$ to $\boldsymbol{G}'''$ we add checks $a'''_{i,j}$, $i \geq 1, j \in [\boldsymbol{\gamma}_i]$ and $b'''_{i,j}$, $i \geq 1, j \in [\boldsymbol{M}_i^+ - \boldsymbol{M}_i^- - \boldsymbol{\gamma}_i]$. Let

$$\mathscr{X} = \left(\bigcup_{i\geq 1}\bigcup_{j=1}^{\boldsymbol{\gamma}_i}\partial a'''_{i,j} \setminus \{x_{n+1}\}\right) \cup \left(\bigcup_{i\geq 1}\bigcup_{j\in[\boldsymbol{M}_i^+ - \boldsymbol{M}_i^- - \boldsymbol{\gamma}_i]}\partial b'''_{i,j}\right),$$

comprise all the variable nodes adjacent to the new checks, except for x_{n+1}. Further, let

$$\mathscr{E}''' = \left\{|\mathscr{X}| = Y + \sum_{i\geq 1}(i-1)\boldsymbol{\gamma}_i\right\},$$

be the event that the variables of $\boldsymbol{G}'$ where the new checks attach are all distinct.

Claim 5.10. We have $\mathbb{P}\left[\mathscr{E}'''|\mathscr{E}' \cap \mathscr{E}''\right] = 1 - o_n(1)$.

Proof. Given $\mathscr{E}'$ there are $\Omega_n(n)$ cavities in total, while the maximum number belonging to any one variable is $O_n(\sqrt{n})$. Further, given $\mathscr{E}''$ we merely pick a bounded number $Y + Y' = O_{\varepsilon,n}(1/\varepsilon)$ of these cavities randomly as neighbors of the new checks. Thus, the probability of hitting the same variable twice is $o_n(1)$. ∎

Claim 5.11. We have $\mathbb{E}\left[\left|\mathrm{nul}(\boldsymbol{A}''') - \mathrm{nul}(\boldsymbol{A}')\right|(1 - \mathbf{1}\mathscr{E} \cap \mathscr{E}' \cap \mathscr{E}'' \cap \mathscr{E}''')\right] = o_{\varepsilon,n}(1)$.

Proof. Clearly $\left|\mathrm{nul}(\boldsymbol{A}''') - \mathrm{nul}(\boldsymbol{A}')\right| \leq X + \boldsymbol{d}_{n+1} + 1$ because going from $\boldsymbol{A}'$ to $\boldsymbol{A}'''$ we add one column and at most $X + \boldsymbol{d}_{n+1}$ new rows. Consequently, as $\mathbb{E}[X^2], \mathbb{E}[\boldsymbol{d}_{n+1}^2] = O_{\varepsilon,n}(1)$, the Cauchy–Schwarz inequality yields

$$\mathbb{E}\left[\left|\mathrm{nul}(\boldsymbol{A}''') - \mathrm{nul}(\boldsymbol{A}')\right|(1 - \mathbf{1}\mathscr{E}'')\right] \leq \mathbb{E}\left[(X + \boldsymbol{d}_{n+1} + 1)^2\right]^{1/2}\left(1 - \mathbb{P}\left[\mathscr{E}''\right]\right)^{1/2} = o_{\varepsilon,n}(1). \tag{5.23}$$

Furthermore, Lemma 5.6 and Claims 5.7–5.10 readily imply that

$$\mathbb{E}\left[\left|\mathrm{nul}(\boldsymbol{A}''') - \mathrm{nul}(\boldsymbol{A}')\right|\mathbf{1}\mathscr{E}'' \setminus \mathscr{E}\right] \leq O_{\varepsilon,n}(\varepsilon^{-1})\exp(-1/\varepsilon^4) = o_{\varepsilon,n}(1), \tag{5.24}$$

$$\mathbb{E}\left[\left|\mathrm{nul}(\boldsymbol{A}''') - \mathrm{nul}(\boldsymbol{A}')\right|\mathbf{1}\mathscr{E}'' \setminus \mathscr{E}'\right], \mathbb{E}\left[\left|\mathrm{nul}(\boldsymbol{A}''') - \mathrm{nul}(\boldsymbol{A}')\right|\mathbf{1}\mathscr{E}'' \cap \mathscr{E}' \setminus \mathscr{E}'''\right] = o_n(1). \tag{5.25}$$

The assertion follows from (5.23)–(5.25). ∎

We obtain $\boldsymbol{G}'''$ by adding checks $a'''_{i,j}$ adjacent to x_{n+1} and $b'''_{i,j}$ not adjacent to x_{n+1}. Recall that $\boldsymbol{\alpha}$ signifies the fraction of frozen cavities. Further, let $\Sigma'' \supset \Sigma'$ be the σ-algebra generated by $\boldsymbol{G}', \boldsymbol{A}', \boldsymbol{M}_-$, $(\boldsymbol{d}_i)_{i\in[n+1]}$, $\boldsymbol{\gamma}, \boldsymbol{M}$ and $\boldsymbol{\Delta}$. The random variable $\boldsymbol{\alpha}$ and the events $\mathscr{E}, \mathscr{E}', \mathscr{E}''$ are Σ''-measurable, but $\mathscr{E}'''$ is not. Indeed, given Σ'' the specific cavities of $\boldsymbol{G}'$ that the new checks $a'''_{i,j}, b'''_{i,j}$ join are still random.

Claim 5.12. On the event $\mathscr{E} \cap \mathscr{E}' \cap \mathscr{E}''$ we have

$$\mathbb{E}\left[\left(\mathrm{nul}(\boldsymbol{A}''') - \mathrm{nul}(\boldsymbol{A}')\right)\mathbf{1}\mathscr{E}''' \,|\Sigma''\right] = o_{\varepsilon,n}(1) + \prod_{i\geq 1}(1 - \boldsymbol{\alpha}^{i-1})^{\boldsymbol{\gamma}_i} - \sum_{i\geq 1}(1 - \boldsymbol{\alpha}^{i-1})\boldsymbol{\gamma}_i - \sum_{i\geq 1}(1 - \boldsymbol{\alpha}^i)(\boldsymbol{M}_i^+ - \boldsymbol{M}_i^- - \boldsymbol{\gamma}_i).$$

Proof. Let

$$\mathscr{A} = \left\{a'''_{ij} : i \geq 1,\ j \in [\boldsymbol{\gamma}_i]\right\},$$

be the set of all the new checks connected to x_{n+1} and let

$$\mathscr{B} = \left\{b'''_{ij} : i \geq 1,\ j \in [\boldsymbol{M}_i^+ - \boldsymbol{M}_i^- - \boldsymbol{\gamma}_i]\right\},$$

be the set of all the new checks not connected to x_{n+1}. Let $\tilde{\boldsymbol{B}}$ be the $\{0,1\}$-matrix whose rows are indexed by $\mathscr{A} \cup \mathscr{B}$ and whose columns are indexed by $V_n = \{x_1, \ldots, x_n\}$ such that for each $a \in \mathscr{A} \cup \mathscr{B}'$ and each $x \in V_n$ the corresponding entry equals one iff $x \in \partial_{\boldsymbol{G}'''} a$. Further, obtain $\boldsymbol{B}$ from $\tilde{\boldsymbol{B}}$ by replacing each one-entry by the entry supplied by χ that represents the respective new edge of the Tanner graph. If the event $\mathscr{E}'''$ occurs, then each column of B contains at most one nonzero entry and each row contains at least one non-zero entry. In effect, $\boldsymbol{B}$ has full rank, that is,

$$\mathrm{rk}(\boldsymbol{B}) = |\mathscr{A} \cup \mathscr{B}| = \sum_{i \geq 1} \boldsymbol{M}_i^+ - \boldsymbol{M}_i^-.$$

Further, let $\boldsymbol{B}_*$ be the matrix obtained from $\boldsymbol{B}$ by replacing all entries in the x-column by zero for every $x \in \mathfrak{F}(\boldsymbol{A}')$. Finally, let $\boldsymbol{C} \in \mathbb{F}^{\mathscr{A} \cup \mathscr{B}}$ be a column vector whose entries $\boldsymbol{C}_a$, $a \in \mathscr{A}$, are the entries from χ representing the edges of the Tanner graph $\boldsymbol{G}'''$ incident with x_{n+1} and whose remaining entries $\boldsymbol{C}_b$, $b \in \mathscr{B}$, are equal to zero.

By construction, on the event $\mathscr{E} \cap \mathscr{E}' \cap \mathscr{E}'' \cap \mathscr{E}'''$ we have

$$\mathrm{nul}\boldsymbol{A}''' = \mathrm{nul}\begin{pmatrix} \boldsymbol{A}' & 0 \\ \boldsymbol{B} & \boldsymbol{C} \end{pmatrix}.$$

Moreover, on $\mathscr{E}'$ the set $\mathscr{X}'''$ of nonzero columns of $\boldsymbol{B}$ has size at most $|\mathscr{X}'''| \leq Y + Y' \leq 2/\varepsilon$, while there are at least $\varepsilon dn/2$ cavities. As a consequence, even though the sequence of cavities that the new checks join are drawn without replacement, this sequence is at total variation distance $o_n(1)$ from a sequence of independent samples from the distribution P. Therefore, on $\mathscr{E} \cap \mathscr{E}' \cap \mathscr{E}''$ the conditional probability given $\mathscr{E}'''$ that $\mathscr{X}'''$ forms a proper relation is bounded by $O_{\varepsilon,n}(\exp(-1/\varepsilon^4))$. Hence, Lemma 2.5 implies that on $\mathscr{E} \cap \mathscr{E}' \cap \mathscr{E}''$,

$$\mathbb{E}\left[\left(\mathrm{nul}(\boldsymbol{A}''') - \mathrm{nul}(\boldsymbol{A}')\right)\mathbf{1}\mathscr{E}'''|\Sigma''\right] = 1 - \mathbb{E}\left[\mathrm{rk}\,(\boldsymbol{B}_*\ \boldsymbol{C})\,|\Sigma''\right] + o_{\varepsilon,n}(1). \tag{5.26}$$

On $\mathscr{E}'''$ the matrix $\boldsymbol{Q} = (\boldsymbol{B}_*\ \boldsymbol{C})$ is a block matrix that decomposes into the $\mathscr{A}$-rows $\boldsymbol{Q}_{\mathscr{A}}$ and the $\mathscr{B}$-rows $\boldsymbol{Q}_{\mathscr{B}}$. Hence, $\mathrm{rk}(\boldsymbol{Q}) = \mathrm{rk}(\boldsymbol{Q}_{\mathscr{A}}) + \mathrm{rk}(\boldsymbol{Q}_{\mathscr{B}})$. To complete the proof, we claim that

$$\mathbb{E}\left[\mathrm{rk}\left(\boldsymbol{Q}_{\mathscr{B}}\right)|\Sigma''\right] = o_n(1) + \sum_{i \geq 1}\left(1 - \boldsymbol{\alpha}^i\right)(\boldsymbol{M}_i^+ - \boldsymbol{M}_i^- - \boldsymbol{\gamma}_i), \tag{5.27}$$

$$\mathbb{E}\left[\mathrm{rk}(\boldsymbol{Q}_{\mathscr{A}})|\Sigma''\right] = o_n(1) + \sum_{i \geq 1}\left(1 - \boldsymbol{\alpha}^{i-1}\right)\boldsymbol{\gamma}_i + 1 - \prod_{i \geq 1}\left(1 - \boldsymbol{\alpha}^{i-1}\right)^{\boldsymbol{\gamma}_i}, \tag{5.28}$$

where, as we recall, $\boldsymbol{\alpha}$ is the probability that a cavity chosen from $p(\cdot)$ is frozen. Indeed, the probability that a $\mathscr{B}$-row of $\boldsymbol{B}$ that contains precisely i nonzero entries gets zeroed out completely in $\boldsymbol{B}_*$ equals $\boldsymbol{\alpha}^i + o_n(1)$ and there are $\boldsymbol{M}_i^+ - \boldsymbol{M}_i^- - \boldsymbol{\gamma}_i$ such rows; hence (5.27).

Similarly, the probability that an $\mathscr{A}$-row of $\boldsymbol{B}$ with $i-1$ nonzero entries gets zeroed out completely in $\boldsymbol{B}_*$ equals $\boldsymbol{\alpha}^{i-1}+o_n(1)$ and there are $\boldsymbol{\gamma}_i$ such rows. Hence, the expected rank of the $\mathscr{A}$-rows of $\boldsymbol{B}_*$ equals $\sum_{i\geq 1}\left(1-\boldsymbol{\alpha}^{i-1}\right)\boldsymbol{\gamma}_i+o_n(1)$, which is the first summand in (5.28). Moreover, the presence of the $\boldsymbol{C}$-column adds one to the rank of $\boldsymbol{Q}_{\mathscr{A}}$ unless not a single one of the $\mathscr{A}$-rows of $\boldsymbol{B}$ gets zero out, which occurs with probability $\prod_{i\geq 1}\left(1-\boldsymbol{\alpha}^{i-1}\right)^{\boldsymbol{\gamma}_i}+o_n(1)$. Hence, we obtain (5.28). Finally, the assertion follows from (5.26)–(5.28). ∎

Proof of Lemma 5.3. Let $\mathfrak{G}=\mathscr{E}\cap\mathscr{E}'\cap\mathscr{E}''\cap\mathscr{E}'''$. Combining Claims 5.9–5.12, we see that

$$\mathbb{E}\left|\mathbb{E}\left[\mathrm{nul}(\boldsymbol{A}''')-\mathrm{nul}(\boldsymbol{A}')|\Sigma''\right]-\left(\prod_{i\geq 1}(1-\boldsymbol{\alpha}^{i-1})^{\boldsymbol{\gamma}_i}-\sum_{i\geq 1}(1-\boldsymbol{\alpha}^{i-1})\boldsymbol{\gamma}_i-\sum_{i\geq 1}(1-\boldsymbol{\alpha}^{i})(\boldsymbol{M}_i^+-\boldsymbol{M}_i^--\boldsymbol{\gamma}_i)\right)\mathbf{1}\mathfrak{G}\right|=o_{\varepsilon,n}(1). \tag{5.29}$$

Since on $\mathfrak{G}$ all degrees i with $\boldsymbol{M}_i^+-\boldsymbol{M}_i^--\boldsymbol{\gamma}_i>0$ are bounded and Chebyshev's inequality shows that $\boldsymbol{M}_i\sim\mathbb{E}[\boldsymbol{M}_i]=\Omega_n(n)$ for any fixed i a.a.s., (5.3) yields $\boldsymbol{M}_i^-=\boldsymbol{M}_i-\boldsymbol{\gamma}_i$ a.a.s. Hence, (5.29) turns into

$$\mathbb{E}\left|\mathbb{E}\left[\mathrm{nul}(\boldsymbol{A}''')-\mathrm{nul}(\boldsymbol{A}')|\Sigma''\right]-\left(\prod_{i\geq 1}(1-\boldsymbol{\alpha}^{i-1})^{\boldsymbol{\gamma}_i}-\sum_{i\geq 1}(1-\boldsymbol{\alpha}^{i-1})\boldsymbol{\gamma}_i-\sum_{i\geq 1}(1-\boldsymbol{\alpha}^{i})\boldsymbol{\Delta}_i\right)\mathbf{1}\mathscr{E}''\right|=o_{\varepsilon,n}(1). \tag{5.30}$$

Further, since $\sum_{i\geq 1}\boldsymbol{\gamma}_i\leq\boldsymbol{d}_{n+1}$ and $\mathbb{E}[\boldsymbol{d}_{n+1}]=O_{\varepsilon,n}(1)$, we obtain

$$\begin{aligned}
&\mathbb{E}\left[\left(\prod_{i\geq 1}(1-\boldsymbol{\alpha}^{i-1})^{\boldsymbol{\gamma}_i}-\sum_{i\geq 1}(1-\boldsymbol{\alpha}^{i-1})\boldsymbol{\gamma}_i\right)\mathbf{1}\mathfrak{G}\right]\\
&=\mathbb{E}\left[\left(\prod_{i\geq 1}(1-\boldsymbol{\alpha}^{i-1})^{\boldsymbol{\gamma}_i}-\sum_{i\geq 1}(1-\boldsymbol{\alpha}^{i-1})\boldsymbol{\gamma}_i\right)\mathbf{1}\mathfrak{G}\cap\left\{\sum_{i\geq 1}\boldsymbol{\gamma}_i\leq\varepsilon^{-1/4}\right\}\right]+o_{\varepsilon,n}(1)\\
&=\mathbb{E}\left[\left(\prod_{i\geq 1}(1-\boldsymbol{\alpha}^{i-1})^{\boldsymbol{\gamma}_i}-\sum_{i\geq 1}(1-\boldsymbol{\alpha}^{i-1})\boldsymbol{\gamma}_i\right)\mathbf{1}\left\{\sum_{i\geq 1}\boldsymbol{\gamma}_i\leq\varepsilon^{-1/4}\right\}\right]+o_{\varepsilon,n}(1)\\
&\quad\text{[by Lemmas 5.6 – 5.7/Claims 5.9 – 5.10]}\\
&=\mathbb{E}\left[\left(\prod_{i\geq 1}(1-\boldsymbol{\alpha}^{i-1})^{\hat{\boldsymbol{\gamma}}_i}-\sum_{i\geq 1}(1-\boldsymbol{\alpha}^{i-1})\hat{\boldsymbol{\gamma}}_i\right)\mathbf{1}\left\{\sum_{i\geq 1}\hat{\boldsymbol{\gamma}}_i\leq\varepsilon^{-1/4}\right\}\right]+o_{\varepsilon,n}(1)\quad\text{[by Lemma 5.8]}\\
&=\mathbb{E}\left[(1-\boldsymbol{\alpha}^{\hat{k}-1})^{\boldsymbol{d}}-\boldsymbol{d}-\boldsymbol{d}\boldsymbol{\alpha}^{\hat{k}-1}\right]+o_{\varepsilon,n}(1)\quad\text{[by the def. of }\hat{\boldsymbol{\gamma}}\text{]}\\
&=\mathbb{E}\left[D(1-K'(\boldsymbol{\alpha})/k)-d-\frac{d}{k}K'(\boldsymbol{\alpha})\right]+o_{\varepsilon,n}(1)\quad\text{[by(1.10)]}.
\end{aligned}\tag{5.31}$$

Similarly, Claim 5.9, Lemma 5.8 and the construction (5.16) of $\hat{\boldsymbol{\Delta}}$ yield

$$\begin{aligned}
\mathbb{E}\left[\left(\sum_{i\geq 1}(1-\boldsymbol{\alpha}^{i})\boldsymbol{\Delta}_i\right)\mathbf{1}\mathscr{E}''\right]&=\mathbb{E}\left[\left(\sum_{i\geq 1}(1-\boldsymbol{\alpha}^{i})\boldsymbol{\Delta}_i\right)\mathbf{1}\left\{\sum_{i\geq 1}\boldsymbol{\Delta}_i\leq\varepsilon^{-1/3}\right\}\right]\\
+o_{\varepsilon,n}(1)&=\mathbb{E}\left[\sum_{i\geq 1}(1-\boldsymbol{\alpha}^{i})\hat{\boldsymbol{\Delta}}_i\right]+o_{\varepsilon,n}(1)
\end{aligned}$$

$$= o_{\varepsilon,n}(1) + (1-\varepsilon)\frac{d}{k}\sum_{i\geq 1}\mathbb{P}\left[\boldsymbol{k}=i\right]\mathbb{E}[1-\boldsymbol{\alpha}^i]$$
$$= o_{\varepsilon,n}(1) + \frac{d}{k} - \frac{d}{k}\mathbb{E}[K(\boldsymbol{\alpha})]. \tag{5.32}$$

Finally, the assertion follows from (5.30), (5.31), and (5.32). ∎

5.4 | Proof of Lemma 5.4

The argument resembles the one from the proof of Lemma 5.3 but the details are considerably more straightforward as we merely add checks to obtain $\boldsymbol{A}''$ from $\boldsymbol{A}'$. As before we consider the events $\mathscr{E}, \mathscr{E}'$ from (5.12) and (67) Moreover, recalling that the total number of new non-zero entries when going from $\boldsymbol{A}'$ to $\boldsymbol{A}''$ is bounded by $\boldsymbol{d}_{n+1}$, we introduce $\mathscr{E}'' = \{\boldsymbol{d}_{n+1} \leq 1/\varepsilon\}$.

Claim 5.13. We have $\mathbb{P}\left[\mathscr{E}''\right] = 1 - O_{\varepsilon,n}(\varepsilon^2)$.

Proof. This follows from the assumption $\mathbb{E}[\boldsymbol{d}_{n+1}^2] = O_{\varepsilon,n}(1)$ and Chebyshev's inequality. ∎

Further, similarly as in the proof of Lemma 5.3 we consider the set

$$\mathscr{X} = \bigcup_{i\geq 1}\bigcup_{j\in[\boldsymbol{M}_i - \boldsymbol{M}_i^-]}\partial_{\boldsymbol{G}''}a''_{i,j},$$

of variable nodes adjacent to the new checks. Let $\mathscr{E}'''$ be the event that none of the variable nodes in $\mathscr{X}$ is connected with the set of new checks by more than one edge.

Claim 5.14. We have $\mathbb{P}\left[\mathscr{E}'''|\mathscr{E}'\cap\mathscr{E}''\right] = 1 - o_n(1)$.

Proof. Given $\mathscr{E}'$ there are $\Omega_n(n)$ cavities in total, with each variable node contributing no more than $O_n(\sqrt{n})$ cavities. Moreover, given $\mathscr{E}''$ we choose $O_{\varepsilon,n}(1/\varepsilon)$ of cavities randomly to attach the new checks. Consequently, the probability of twice choosing a cavity with the same underlying variable is $o_n(1)$. ∎

Claim 5.15. We have $\mathbb{E}\left[\left|\mathrm{nul}(\boldsymbol{A}'') - \mathrm{nul}(\boldsymbol{A}')\right|(1 - \mathbf{1}\mathscr{E}\cap\mathscr{E}'\cap\mathscr{E}''\cap\mathscr{E}''')\right] = o_{\varepsilon,n}(1)$.

Proof. We have $\left|\mathrm{nul}(\boldsymbol{A}'') - \mathrm{nul}(\boldsymbol{A}')\right| \leq \boldsymbol{d}_{n+1}$ as we add at most $\boldsymbol{d}_{n+1}$ rows. Because $\mathbb{E}[\boldsymbol{d}_{n+1}] = O_{\varepsilon,n}(1)$, Claim (5.13) and the Cauchy–Schwarz inequality yield

$$\mathbb{E}\left[\left|\mathrm{nul}(\boldsymbol{A}'') - \mathrm{nul}(\boldsymbol{A}')\right|(1 - \mathbf{1}\mathscr{E}'')\right] \leq \mathbb{E}\left[\boldsymbol{d}_{n+1}^2\right]^{1/2}(1 - \mathbb{P}\left[\mathscr{E}\right])^{1/2} = o_{\varepsilon,n}(1). \tag{5.33}$$

Moreover, Lemma 5.6, Lemma 5.7, and Claim 5.14 show that

$$\mathbb{E}\left[\left|\mathrm{nul}(\boldsymbol{A}'') - \mathrm{nul}(\boldsymbol{A}')\right|\mathbf{1}\mathscr{E}''\setminus\mathscr{E}\right], \mathbb{E}\left[\left|\mathrm{nul}(\boldsymbol{A}'') - \mathrm{nul}(\boldsymbol{A}')\right|\mathbf{1}\mathscr{E}''\setminus\mathscr{E}'\right],$$
$$\mathbb{E}\left[\left|\mathrm{nul}(\boldsymbol{A}'') - \mathrm{nul}(\boldsymbol{A}')\right|\mathbf{1}\mathscr{E}''\setminus\mathscr{E}'''\right] = o_{\varepsilon,n}(1). \tag{5.34}$$

The assertion follows from (5.33) and(5.34). ∎

The matrix $\boldsymbol{A}''$ results from $\boldsymbol{A}'$ by adding checks $a''_{i,j}$, $i \geq 1, j \in [\boldsymbol{M}_i - \boldsymbol{M}_i^-]$ that are connected to random cavities of $\boldsymbol{A}'$. Moreover, as before let $\Sigma'' \supset \Sigma'$ be the σ-algebra generated by $\boldsymbol{G}'$, $\boldsymbol{A}'$, $\boldsymbol{M}_-$, $(\boldsymbol{d}_i)_{i\in[n+1]}$, $\boldsymbol{\gamma}, \boldsymbol{M}$ and $\boldsymbol{\Delta}$. Then $\mathscr{E}, \mathscr{E}', \mathscr{E}''$ are Σ''-measurable, but $\mathscr{E}'''$ is not.

Claim 5.16. On the event $\mathscr{E} \cap \mathscr{E}' \cap \mathscr{E}''$ we have $\mathbb{E}\left[(\mathrm{nul}(\boldsymbol{A}'') - \mathrm{nul}(\boldsymbol{A}'))\mathbf{1}\mathscr{E}'''|\Sigma''\right] = o_{\varepsilon,n}(1) - \sum_{i\geq 1}(1-\boldsymbol{\alpha}^i)(\boldsymbol{M}_i - \boldsymbol{M}_i^-)$.

Proof. Let $\mathscr{A}$ be the set of all the new checks $a''_{i,j}$, $i \geq 1, j \in [\boldsymbol{M}_i - \boldsymbol{M}_i^-]$. Let $\tilde{\boldsymbol{B}}$ be the $\{0,1\}$-matrix whose rows are indexed by $\mathscr{A}$ and whose columns are indexed by $V_n = \{x_1, \ldots, x_n\}$ such that for each $a \in \mathscr{A}$ and each $x \in V_n$ the corresponding entry equals one iff $x \in \partial_{\boldsymbol{G}''} a$. Further, obtain $\boldsymbol{B}$ by substituting each one-entry of $\tilde{\boldsymbol{B}}$ by the appropriate nonzero field element from χ. If $\mathscr{E}'''$ occurs, then $\boldsymbol{B}$ has rank $\mathrm{rk}(\boldsymbol{B}) = |\mathscr{A}| = \sum_{i\geq 1} \boldsymbol{M}_i^+ - \boldsymbol{M}_i$, because no column contains two nonzero entries and each row contains at least one nonzero entry. Further, let $\boldsymbol{B}_*$ be the matrix obtained from $\boldsymbol{B}$ by replacing all entries in the x-column by zero if $x \in \mathfrak{F}(\boldsymbol{A}')$ is frozen to zero in $\boldsymbol{A}'$.

On the event $\mathscr{E} \cap \mathscr{E}' \cap \mathscr{E}'' \cap \mathscr{E}'''$ we have

$$\mathrm{nul}\boldsymbol{A}'' = \mathrm{nul}\begin{pmatrix}\boldsymbol{A}'\\ \boldsymbol{B}\end{pmatrix}. \tag{5.35}$$

Moreover, on $\mathscr{E}' \cap \mathscr{E}''$ the set $\mathscr{X}''$ of non-zero columns of $\boldsymbol{B}$ has size at most $|\mathscr{X}''| \leq \boldsymbol{d}_{n+1} \leq 1/\varepsilon$, while there are at least $\varepsilon dn/2$ cavities. Hence, on $\mathscr{E} \cap \mathscr{E}' \cap \mathscr{E}'' \cap \mathscr{E}'''$ the probability that $\mathscr{X}''$ forms a proper relation is bounded by $\exp(-1/\varepsilon^4)$. Therefore, Lemma 2.5 implies that

$$\mathbb{E}\left[\left(\mathrm{nul}(\boldsymbol{A}'') - \mathrm{nul}(\boldsymbol{A}')\right)\mathbf{1}\mathscr{E}'''|\Sigma''\right] = o_{\varepsilon,n}(1) - \mathbb{E}\left[\mathrm{rk}\left(\boldsymbol{B}_*\right)|\Sigma''\right]. \tag{5.36}$$

Further, since an $\boldsymbol{\alpha}$-fraction of cavities are frozen, a row of $\boldsymbol{B}$ with i nonzero entries gets zeroed out completely in $\boldsymbol{B}_*$ with probability $\boldsymbol{\alpha}^i + o_n(1)$. Consequently,

$$\mathbb{E}\left[\mathrm{rk}\left(\boldsymbol{B}_*\right)|\Sigma''\right] = o_{\varepsilon,n}(1) + \sum_{i\geq 1}\left(1-\boldsymbol{\alpha}^i\right)(\boldsymbol{M}_i - \boldsymbol{M}_i^-). \tag{5.37}$$

Finally, the assertion follows from (5.36) and (5.37). ∎

Proof of Lemma 5.4. Let $\mathfrak{E} = \mathscr{E} \cap \mathscr{E}' \cap \mathscr{E}'' \cap \mathscr{E}'''$. Combining Claims 5.15–5.16, we obtain

$$\mathbb{E}\left|\mathbb{E}[\mathrm{nul}(\boldsymbol{A}'') - \mathrm{nul}(\boldsymbol{A}')|\Sigma''] + \left(\sum_{i\geq 1}(1-\boldsymbol{\alpha}^i)(\boldsymbol{M}_i - \boldsymbol{M}_i^-)\right)\mathbf{1}\mathfrak{E}\right| = o_{\varepsilon,n}(1). \tag{5.38}$$

Since on $\mathfrak{E}$ all degrees i with $\boldsymbol{M}_i^+ - \boldsymbol{M}_i^- > 0$ are bounded a.a.s. and $\boldsymbol{M}_i^- = \Omega_n(n)$ a.a.s., we conclude that $\boldsymbol{M}_i - \boldsymbol{M}_i^- = \boldsymbol{\gamma}_i$ for all $i \geq 1$ a.a.s. Hence, (5.38) turns into

$$\mathbb{E}\left|\mathbb{E}[\mathrm{nul}(\boldsymbol{A}'') - \mathrm{nul}(\boldsymbol{A}')|\Sigma''] + \left(\sum_{i\geq 1}(1-\boldsymbol{\alpha}^i)\boldsymbol{\gamma}_i\right)\mathbf{1}\mathfrak{E}\right| = o_{\varepsilon,n}(1). \tag{5.39}$$

Further, because $\sum_{i\geq 1}\boldsymbol{\gamma}_i \leq \boldsymbol{d}_{n+1}$ and $\mathbb{E}[\boldsymbol{d}_{n+1}] = O_{\varepsilon,n}(1)$,

$$\mathbb{E}\left[\left(\sum_{i\geq 1}(1-\boldsymbol{\alpha}^i)\boldsymbol{\gamma}_i\right)\mathbf{1}\mathfrak{E}\right]$$

$$= \mathbb{E}\left[\left(\sum_{i\geq 1}(1-\boldsymbol{\alpha}^i)\boldsymbol{\gamma}_i\right)\mathbf{1}\left\{\sum_{i\geq 1}\boldsymbol{\gamma}_i \leq \varepsilon^{-1/4}\right\}\right] + o_{\varepsilon,n}(1)[\text{by Claim 5.13}]$$

$$= \mathbb{E}\left[\left(\sum_{i\geq 1}(1-\boldsymbol{\alpha}^i)\hat{\boldsymbol{\gamma}}_i\right)\mathbf{1}\left\{\sum_{i\geq 1}\hat{\boldsymbol{\gamma}}_i \leq \varepsilon^{-1/4}\right\}\right] + o_{\varepsilon,n}(1) \quad [\text{by Lemma 5.8}]$$

$$= d\mathbb{E}[1-\boldsymbol{\alpha}^{\hat{\boldsymbol{k}}}] + o_\varepsilon(1) = -d\mathbb{E}[\boldsymbol{\alpha}K'(\boldsymbol{\alpha})]/k + d + o_{\varepsilon,n}(1) \quad [\text{by}(1.10)]. \tag{5.40}$$

The assertion follows from (5.39) and (5.40). ∎

5.5 | Proof of Lemma 5.2

Once more we break the proof down into a few relatively simple steps.

Claim 5.17. We have $\mathbb{E}[\mathrm{nul}(\boldsymbol{A}'')] = \mathbb{E}[\mathrm{nul}(\boldsymbol{A}_{n,\boldsymbol{M}})] + o_n(1)$.

Proof. The choice of the random variables in (5.1) and Lemma 1.11 ensure that the event $\mathscr{E} = \left\{\sum_{i\geq 1} i\boldsymbol{M}_i \leq dn/k\right\}$ has probability $1 - o_n(1/n)$. Further, given $\mathscr{E}$ the random variables $\mathrm{nul}(\boldsymbol{A}'')$ and $\mathrm{nul}(\boldsymbol{A}_{n,\boldsymbol{M}})$ are identically distributed by the principle of deferred decisions. Because the nullity of either matrix is bounded by n deterministically, the claim follows. ∎

To compare $\mathrm{nul}(\boldsymbol{A}''')$ and $\mathrm{nul}(\boldsymbol{A}_{n+1,\boldsymbol{M}^+})$ we consider the event

$$\mathscr{E}^+ = \left\{\frac{dn}{2k} \leq \sum_{i\geq 1} i\boldsymbol{M}_i^+ \leq \sum_{i=1}^{n} \boldsymbol{d}_i, \forall i \geq n/\ln^9 n : \boldsymbol{M}_i^+ = 0\right\}.$$

Claim 5.18. We have $\mathbb{P}\left[\mathscr{E}^+\right] = 1 - o_n(1/n)$.

Proof. This follows from the definition (5.2) of the random variables $\boldsymbol{M}_i^+$ and Lemma 1.11. ∎

Further, consider the event

$$\mathscr{W} = \left\{\boldsymbol{d}_{n+1} \leq \ln n, \sum_{i\geq 1} i(\boldsymbol{\Delta}_i + \boldsymbol{\gamma}_i) < \ln^4 n\right\}.$$

Claim 5.19. We have $\mathbb{P}[\mathscr{W}] = 1 - o_n(1)$.

Proof. This follows from the assumption that $\mathbb{E}[\boldsymbol{d}^2], \mathbb{E}[\boldsymbol{k}^2]$ are bounded. ∎

Moreover, let $\mathscr{U}$ be the event that x_{n+1} does not partake in any multi-edges of $\boldsymbol{G}_{n,\boldsymbol{M}^+}$.

Claim 5.20. We have $\mathbb{P}\left[\mathscr{U} \mid \mathscr{W} \cap \mathscr{E}^+\right] = 1 - o_n(\ln^{-6} n)$.

Proof. Given $\mathscr{W} \cap \mathscr{E}^+$ variable node x_{n+1} has target degree at most $\ln n$ and all check degrees are bounded by $n/\ln^9 n$. Hence, the probability that x_{n+1} joins the same check twice is $O_n(\ln^{-7} n)$. ∎

The next claim shows that $\mathrm{nul}(\boldsymbol{A}''')$, $\mathrm{nul}(\boldsymbol{A}_{n+1,\boldsymbol{M}^+})$ can be coupled identically on the 'bulk' event $\mathscr{E}^+ \cap \mathscr{U} \cap \mathscr{W}$.

Claim 5.21. Given $\mathscr{E}^+ \cap \mathscr{U} \cap \mathscr{W}$ the random variables $\mathrm{nul}(\boldsymbol{A}''')$ and $\mathrm{nul}(\boldsymbol{A}_{n+1,\boldsymbol{M}^+})$ are identically distributed and thus

$$\mathbb{E}\left[\left(\mathrm{nul}(\boldsymbol{A}''') - \mathrm{nul}(\boldsymbol{A}_{n+1,\boldsymbol{M}^+})\right)\mathbf{1}\mathscr{U} \cap \mathscr{W} \cap \mathscr{E}^+\right] = 0. \tag{5.41}$$

Proof. By construction, on $\mathscr{E}^+ \cap \mathscr{U} \cap \mathscr{W}$ the random matrices $\boldsymbol{A}'''$ and $\boldsymbol{A}_{n+1,\boldsymbol{M}^+}$ are identically distributed, and hence so are their nullities. ∎

In light of Claims 5.18 and 5.21 we are left to bound the difference of the nullities on $\mathscr{E}^+\setminus(\mathscr{U}\cap\mathscr{W})$.

Claim 5.22. There is a coupling of $\boldsymbol{A}_{n+1,\boldsymbol{M}^+}$ and $\boldsymbol{A}'''$ on $\mathscr{E}^+$ such that $\left|\mathrm{nul}(\boldsymbol{A}''') - \mathrm{nul}(\boldsymbol{A}_{n+1,\boldsymbol{M}^+})\right| \leq 2\sum_{i\geq 1} i(\boldsymbol{\Delta}_i + \boldsymbol{\gamma}_i)$.

Proof. We estimate the number of edges of the Tanner graph $\boldsymbol{G}_{n+1,\boldsymbol{M}^+}$ incident with the checks $a_{i,j}$, $\boldsymbol{M}_i^- < j \leq \boldsymbol{M}_i^+$ or the new variable x_{n+1} of $\boldsymbol{G}_{n+1,\boldsymbol{M}^+}$. By construction, there are at most $\sum_{i\geq 1} i(\boldsymbol{\Delta}_i + \boldsymbol{\gamma}_i)$ such edges. Similarly, there are no more than $\sum_{i\geq 1} i(\boldsymbol{\Delta}_i + \boldsymbol{\gamma}_i)$ edges incident with the new checks $a'''_{i,j}$, $b'''_{i,j}$ added to $\boldsymbol{A}'$ to obtain $\boldsymbol{A}'''$. By the principle of deferred decisions on $\mathscr{E}''$ we can couple the Tanner graphs of $\boldsymbol{A}'''$ and $\boldsymbol{A}_{n+1,\boldsymbol{M}^+}$ such that they coincide on all the edges that join variables $x_1, \ldots, x_n$ and checks $a_{i,j}$, $j \leq \boldsymbol{M}_i^-$, and hence the matrices themselves so that they coincide on all the corresponding matrix entries. Consequently, $\boldsymbol{A}'''$ and $\boldsymbol{A}_{n+1,\boldsymbol{M}^+}$ differ in no more than $2\sum_{i\geq 1} i(\boldsymbol{\Delta}_i + \boldsymbol{\gamma}_i)$ entries. ∎

We proceed to bound the difference of the nullities on $\mathscr{E}^+ \setminus \mathscr{W}$.

Claim 5.23. We have $\mathbb{E}\left[\sum_{i\geq 1} i(\boldsymbol{\Delta}_i + \boldsymbol{\gamma}_i)\mathbf{1}\mathscr{E}^+ \setminus \mathscr{W}\right] = o_n(1)$.

Proof. The event $\mathscr{E}^+ \setminus \mathscr{W}$ is contained in the union of the three events

$$\mathscr{Q}_1 = \mathscr{E}^+ \cap \left\{\exists i > \log n : \boldsymbol{\gamma}_i > 0\right\}, \quad \mathscr{Q}_2 = \mathscr{E}^+ \cap \{\boldsymbol{d}_{n+1} > \log n\} \setminus \mathscr{Q}_1,$$

$$\mathscr{Q}_3 = \mathscr{E}^+ \cap \left\{\sum_{i\geq 1} i\boldsymbol{\Delta}_i > \ln^3 n\right\} \setminus (\mathscr{Q}_1 \cup \mathscr{Q}_2).$$

To bound the contribution of $\mathscr{Q}_1$, consider $\boldsymbol{m}^+_{\varepsilon,n} = \sum_{i\geq 1} \boldsymbol{M}_i^+ \sim \mathrm{Po}((1-\varepsilon)d(n+1)/k)$. We claim that, with the copies $(\boldsymbol{k}_i)_{i\geq 1}$ of $\boldsymbol{k}$ independent of everything else,

$$\mathbb{E}\left[\sum_{i\geq 1} i\boldsymbol{\gamma}_i\mathbf{1}\mathscr{Q}_1\right] \leq O_n(1/n)\cdot\left(1 + \mathbb{E}\left[\sum_{i=1}^{\boldsymbol{m}^+_{\varepsilon,n}}\mathbf{1}\{\boldsymbol{k}_i \geq \log n\}\boldsymbol{k}_i^2\boldsymbol{d}_{n+1}\right]\right) =$$

$$O_n(1)\cdot\mathbb{P}\left[\boldsymbol{k} \geq \log n\right] + O_n(1/n) = O_n(\log^{-2} n). \tag{5.42}$$

Indeed, the last equality sign follows from the first because $\mathbb{E}[\boldsymbol{k}^2] = O_n(1)$ and the first equality sign follows because $\boldsymbol{m}^+_{\varepsilon,n}$ is independent of $\boldsymbol{d}_{n+1}$ and the $\boldsymbol{k}_i$. Further, to obtain the first inequality we consider the $\boldsymbol{m}^+_{\varepsilon,n}$ checks one by one. The degree of the ith check is distributed as $\boldsymbol{k}_i$. We discard it unless $\boldsymbol{k}_i \geq \log n$. But if $\boldsymbol{k}_i \geq \log n$, then the probability that $\boldsymbol{k}_i$ is adjacent to x_{n+1} is bounded by $O_n(\boldsymbol{k}_i\boldsymbol{d}_{n+1}/\sum_{h=1}^{n+1}\boldsymbol{d}_h)$ and $\sum_{h=1}^{n+1}\boldsymbol{d}_h \geq n$. Thus, we obtain (5.42). Further, we observe that (5.42) yields $\mathbb{P}[\mathscr{Q}_1] \leq \mathbb{E}\sum_{i\geq 1} i\boldsymbol{\gamma}_i\mathbf{1}\mathscr{Q}_1 = O_n(\log^{-2} n)$. Hence, as $\mathbb{E}\sum_{i\geq 1} i\boldsymbol{\Delta}_i = O_n(1)$ we obtain

$$\mathbb{E}\left[\sum_{i\geq 1} i\boldsymbol{\Delta}_i \mathbf{1}\mathcal{Q}_1\right] \leq \mathbb{P}\left[\mathcal{Q}_1\right]\log n + \mathbb{E}\left[\sum_{i\geq 1} i\boldsymbol{\Delta}_i \mathbf{1}\left\{\sum_{i\geq 1} i\boldsymbol{\Delta}_i \geq \log n\right\}\right] = o_n(1). \tag{5.43}$$

Combining (5.42) and (5.43), we conclude that

$$\mathbb{E}\left[\sum_{i\geq 1} i(\boldsymbol{\Delta}_i + \boldsymbol{\gamma}_i)\mathbf{1}\mathcal{Q}_1\right] = o_n(1). \tag{5.44}$$

Regarding $\mathcal{Q}_2$, we deduce from the bound $\mathbb{E}[\boldsymbol{d}_{n+1}^r] = O_n(1)$ for an $r > 2$ that

$$\mathbb{E}\left[\sum_{i\geq 1} i\boldsymbol{\gamma}_i \mathbf{1}\mathcal{Q}_2\right] \leq O_n(\log n)\mathbb{E}\left[\boldsymbol{d}_{n+1}\mathbf{1}\{\boldsymbol{d}_{n+1} > \log n\}\right] = o_n(1). \tag{5.45}$$

Moreover, since the $\boldsymbol{\Delta}_i$ are independent of $\boldsymbol{d}_{n+1}$ and $\mathbb{E}\sum_{i\geq 1} i\boldsymbol{\Delta}_i = O_n(1)$, we obtain $\mathbb{E}\left[\sum_{i\geq 1} i\boldsymbol{\Delta}_i \mathbf{1}\mathcal{Q}_2\right] = o_n(1)$. Hence, (5.45) yields

$$\mathbb{E}\left[\sum_{i\geq 1} i(\boldsymbol{\Delta}_i + \boldsymbol{\gamma}_i)\mathbf{1}\mathcal{Q}_2\right] = o_n(1). \tag{5.46}$$

Moving on to $\mathcal{Q}_3$ and recalling the definition (5.2) of $\boldsymbol{\Delta}$, we find

$$\mathbb{P}\left[\mathcal{Q}_3\right] \leq \mathbb{E}\left[\sum_{i\geq 1} i\boldsymbol{\Delta}_i\right]\ln^{-3} n = O_n(\mathbb{E}[\boldsymbol{k}^2]\ln^{-3} n) = o_n(\log^{-2} n). \tag{5.47}$$

Moreover, on $\mathcal{Q}_3$ we have $\sum_{i\geq 1} i\boldsymbol{\gamma}_i \leq \log^2 n$ because $\boldsymbol{d}_{n+1} \leq \log n$ and $\boldsymbol{\gamma}_i = 0$ for all $i \geq \log n$. Consequently, since the $\boldsymbol{\Delta}_i$ are mutually independent and $\sum_{i\geq 1}\mathbb{E}[i\boldsymbol{\Delta}_i] = O_n(1)$, (5.47) yields

$$\mathbb{E}\left[\sum_{i\geq 1} i(\boldsymbol{\Delta}_i + \boldsymbol{\gamma}_i)\mathbf{1}\mathcal{Q}_3\right] \leq o_n(1) + 4\mathbb{E}\left[\sum_{i\geq 1} i\boldsymbol{\Delta}_i \mathbf{1}\left\{\sum_{i\geq 1} i\boldsymbol{\Delta}_i \geq \ln^3 n\right\}\right] = o_n(1). \tag{5.48}$$

Finally, the assertion follows from (5.44), (5.46), and (5.48). ∎

Proof of Lemma 5.2. The first assertion concerning $\boldsymbol{A}''$ and $\boldsymbol{A}_{n,\boldsymbol{M}}$ follows from Claim 5.17.

Concerning $\boldsymbol{A}'''$ and $\boldsymbol{A}_{n+1,\boldsymbol{M}^+}$, Claim 5.18 shows that it suffices to couple $\mathrm{nul}(\boldsymbol{A}''')\mathbf{1}\mathcal{E}^+$ and $\mathrm{nul}(\boldsymbol{A}_{n+1,\boldsymbol{M}^+})\mathbf{1}\mathcal{E}^+$, because both random variables are bounded by $n+1$. Indeed, thanks to Claim 5.21 we merely need to couple $\mathrm{nul}(\boldsymbol{A}''')\mathbf{1}\mathcal{E}^+ \setminus (\mathcal{U}\cap\mathcal{W})$ and $\mathrm{nul}(\boldsymbol{A}_{n+1,\boldsymbol{M}^+})\mathbf{1}\mathcal{E}^+ \setminus (\mathcal{U}\cap\mathcal{W})$, and Claim (5.22) supplies a coupling such that

$$\left|\mathrm{nul}(\boldsymbol{A}''')\mathbf{1}\mathcal{E}^+ - \mathrm{nul}(\boldsymbol{A}_{n+1,\boldsymbol{M}^+})\right|\mathbf{1}\mathcal{E}^+ \setminus (\mathcal{U}\cap\mathcal{W}) \leq 2\sum_{i\geq 1} i(\boldsymbol{\Delta}_i + \boldsymbol{\gamma}_i)\mathbf{1}\mathcal{E}^+ \setminus (\mathcal{U}\cap\mathcal{W}). \tag{5.49}$$

Hence, it suffices to show that

$$\mathbb{E}\left[\sum_{i\geq 1} i(\boldsymbol{\Delta}_i + \boldsymbol{\gamma}_i)\mathbf{1}\mathcal{E}^+ \setminus (\mathcal{U}\cap\mathcal{W})\right] = o_n(1). \tag{5.50}$$

Indeed, in light of Claim 5.23 we merely need to estimate $\sum_{i\geq 1} i(\boldsymbol{\Delta}_i + \boldsymbol{\gamma}_i)\mathbf{1}\{\mathscr{E}^+ \cap \mathscr{W} \setminus \mathscr{U}\}$. But since on $\mathscr{E}^+ \cap \mathscr{W}$ we have $\sum_{i\geq 1} i(\boldsymbol{\Delta}_i + \boldsymbol{\gamma}_i) \leq \ln^4 n$, Claim 5.20 yields

$$\mathbb{E}\left[\sum_{i\geq 1} i(\boldsymbol{\Delta}_i + \boldsymbol{\gamma}_i)\mathbf{1}\{\mathscr{E}^+ \cap \mathscr{W} \setminus \mathscr{U}\}\right] \leq \left(1 - \mathbb{P}\left[\mathscr{U} \mid \mathscr{E}^+ \cap \mathscr{W}\right]\right)\ln^4 n = o_n(1). \tag{5.51}$$

Finally, the assertion follows from Claim 5.23 and (5.49)–(5.51). ∎

6 | PROOF OF THEOREM 1.2

We describe how to extend the proof of [47] to $\boldsymbol{G}$. First, we will work on $G = \boldsymbol{G}_{0,n}$ (defined in Section 2.2), the configuration model for $\boldsymbol{G}$. By Lemma 4.3, properties that holds with probability $1 - o_n(1)$ for G also hold with probability $1 - o_n(1)$ for $\boldsymbol{G}$. Second, using the terminology in [47], variable nodes in G are called vertices, and each check node corresponds to a hyperedge in the following sense: if f_a is a check node adjacent with variable nodes $\{v_{a_1}, \ldots, v_{a_h}\}$ for some $h \geq 1$, then the set of vertices $\{v_{a_1}, \ldots, v_{a_h}\}$ is called a hyperedge. A check node with size 0 corresponds to an isolated hyperedge with size 0, i.e. this hyperedge does not contain any vertex.

Suppose $\mathbb{P}(\boldsymbol{d} \geq 2) > 0$ as otherwise the theorem holds trivially by Remark 1.3(b). We first prove Theorem 1.2 in the case $\boldsymbol{k} \geq 1$. Consider the parallel stripping process where all vertices of degree less than 2 are deleted in each step, together with the hyperedges (if any) incident with them. Take a random vertex $v \in [n]$. Let λ_t be the probability that v survives after t iterations of the stripping process. It is easy to see that λ_t is monotonically non-increasing and thus $\lambda = \lim_{t\to\infty} \lambda_t$ exists. For any vertex $u \in [n]$, let $\partial^j(u)$ denote the set of vertices of distance j from u. Recall that there exists a constant $\sigma > 0$ such that $\mathbb{E}\boldsymbol{d}^{2+\sigma} < \infty$ and $\mathbb{E}\boldsymbol{k}^{2+\sigma} < \infty$ by our assumptions on $\boldsymbol{d}$ and $\boldsymbol{k}$. We claim that

Claim 6.1. With high probability, the maximum degree and the maximum size of hyperedges in G are at most $(n\log n)^{1/(2+\sigma)}$, and for every $u \in [n]$ and for all fixed R, $|\cup_{j\leq R} \partial^j(u)| = O_n(n^{1/(2+\sigma)}\log^2 n)$.

Let H_t be the subgraph of G obtained after t iterations of the parallel stripping process. Consider Doob's martingale $(\mathbb{E}(H_t|e_1, \ldots, e_j))_{0\leq j\leq m}$ where random hyperedges are added in the order $e_1, \ldots, e_m$ using the configuration model, and m denotes the number of hyperedges in G. By Claim 6.1, swapping two clones in the configuration model would affect H_t by $O_n(n^{1/(2+\sigma)}\log^2 n)$, as each altered hyperedge can only affect the vertices (if surviving the first tth iteration or not) within its t-neighborhood. Standard concentration arguments (see, for instance, the proof of [, theorem 2.19]) based on Azuma's inequality (with Lipschitz constant $Cn^{1/(2+\sigma)}\log^2 n$ for some fixed $C > 0$) produce that $\|H_t| - \lambda_t n| = O_n(n^{(4+\sigma)/(4+2\sigma)}\log^3 n) = o_n(n)$. Next we deduce an expression for λ_t. Consider a random hypertree T iteratively built as follows. The root of T is v, which is incident to d_v hyperedges of size $\boldsymbol{k}_1, \ldots, \boldsymbol{k}_{d_v}$ where the $\boldsymbol{k}_i$s are i.i.d. copies of $\hat{\boldsymbol{k}}$ where

$$\mathbb{P}(\hat{\boldsymbol{k}} = j) = \frac{j\mathbb{P}(\boldsymbol{k} = j)}{k}. \tag{6.1}$$

Then the ith hyperedge is incident to other $\boldsymbol{k}_i - 1$ vertices (other than v) whose degrees are i.i.d. copies of $\hat{\boldsymbol{d}}$, where

$$\mathbb{P}(\hat{\boldsymbol{d}} = j) = \frac{j\mathbb{P}(\boldsymbol{d} = j)}{d}. \tag{6.2}$$

This builds the first neighborhood of v in T. Iteratively we can build the r-neighborhood of v in T for any fixed r. It follows from the following claim that the r-neighborhood of v in G converges in distribution to the r-neighborhood of T, as $n \to \infty$, for any fixed $r \geq 1$. This is because when uniformly picking a random variable clone (or check clone), the degree of the corresponding variable node (or check node) has the distribution in (6.2) (or (6.1)). Let S be a set of vertices in G. We say S induces a cycle if there is a closed walk $x_0x_1 \ldots x_\ell = x_0$ such that all $x_i \in S$, and every pair of consecutive vertices in the walk are contained in a hyperedge in G.

Claim 6.2. With high probability, for all fixed $R \geq 1$, $\cup_{j\leq R}\partial^j(v)$ induces no cycles.

(The proofs of Claims 6.1 and 6.2 can be found at the end of this section.) If v survives t iterations of the stripping process then at least two hyperedges incident with v survives after t iterations of the stripping process. On the other hand, let x be a hyperedge of size at least 1 and let u be a vertex incident with x. Let ρ_t denote the probability that u is incident with at least one hyperedge other than x which survives after t iterations of the stripping process. We will deduce a recursion for ρ_t and then deduce λ_t from ρ_t. Note that the degree of u follows the distribution from (6.2). Then, ignoring an $o_n(1)$ error accounting for the probability of the complement of the events in Claims 6.1 and 6.2:

$$\rho_0 = 1,$$

and

$$\begin{aligned}
\rho_{t+1} &= \sum_{j\geq 2}\frac{j\mathbb{P}(\boldsymbol{d}=j)}{d}\sum_{S\subseteq[j-1],|S|\geq 1}\sum_{k_1,\ldots,k_{j-1}\geq 1}\prod_{i=1}^{j-1}\mathbb{P}(\hat{\boldsymbol{k}}_i=k_i)\prod_{i\in S}\rho_t^{k_i-1}\prod_{i\in[j-1]\setminus S}(1-\rho_t)^{k_i-1}\\
&= \sum_{j\geq 2}\frac{j\mathbb{P}(\boldsymbol{d}=j)}{d}\sum_{h\geq 1}\binom{j-1}{h}\left(\sum_{k'\geq 1}\mathbb{P}(\hat{\boldsymbol{k}}=k')\rho_t^{k'-1}\right)^h\left(\sum_{k'\geq 1}\mathbb{P}(\hat{\boldsymbol{k}}=k')(1-\rho_t)^{k'-1}\right)^{j-1-h}\\
&= \sum_{j\geq 2}\frac{j\mathbb{P}(\boldsymbol{d}=j)}{d}\sum_{h\geq 1}\binom{j-1}{h}\left(\frac{K'(\rho_t)}{k}\right)^h\left(1-\frac{K'(\rho_t)}{k}\right)^{j-1-h}\\
&= \sum_{j\geq 2}\frac{j\mathbb{P}(\boldsymbol{d}=j)}{d}\left(1-\left(1-\frac{K'(\rho_t)}{k}\right)^{j-1}\right) = 1-\frac{D'(1-\frac{K'(\rho_t)}{k})}{d},
\end{aligned}$$

noting that

$$\mathbb{E}\rho^{\hat{k}-1} = \sum_{k'\geq 1}\mathbb{P}(\hat{\boldsymbol{k}}=k')\rho^{k'-1} = \sum_{j\geq 1}\frac{j\mathbb{P}(\boldsymbol{k}=j)}{k}\rho^{j-1} = \frac{K'(\rho)}{k}.$$

Consequently,

$$\begin{aligned}
\lambda_t &= \sum_{j\geq 2}\mathbb{P}(\boldsymbol{d}=j)\sum_{h\geq 2}\binom{j}{h}\left(\sum_{k'\geq 1}\mathbb{P}(\hat{\boldsymbol{k}}=k')\rho_t^{k'-1}\right)^h\left(1-\sum_{k'\geq 1}\mathbb{P}(\hat{\boldsymbol{k}}=k')\rho_t^{k'-1}\right)^{j-h}\\
&= \sum_{j\geq 2}\mathbb{P}(\boldsymbol{d}=j)\sum_{h\geq 2}\binom{j}{h}\left(\mathbb{E}\rho_t^{\hat{\boldsymbol{k}}-1}\right)^h\left(1-\mathbb{E}\rho_t^{\hat{\boldsymbol{k}}-1}\right)^{j-h}\\
&= \sum_{j\geq 2}\mathbb{P}(\boldsymbol{d}=j)\left(1-\left(1-\frac{K'(\rho_t)}{k}\right)^j-j\frac{K'(\rho_t)}{k}\left(1-\frac{K'(\rho_t)}{k}\right)^{j-1}\right)
\end{aligned}$$

$$= \mathbb{P}(\boldsymbol{d} \geq 2) - \left(D\left(1 - \frac{K'(\rho_t)}{k}\right) - \mathbb{P}(\boldsymbol{d} = 0) - \mathbb{P}(\boldsymbol{d} = 1)\left(1 - \frac{K'(\rho_t)}{k}\right)\right)$$
$$- \left(\frac{K'(\rho_t)}{k} D\boldsymbol{\prime}\left(1 - \frac{K'(\rho_t)}{k}\right) - \mathbb{P}(\boldsymbol{d} = 1)\frac{K'(\rho_t)}{k}\right)$$
$$= 1 - D\left(1 - \frac{K'(\rho_t)}{k}\right) - \frac{K'(\rho_t)}{k} D\boldsymbol{\prime}\left(1 - \frac{K'(\rho_t)}{k}\right).$$

Let $g(x) = 1 - \frac{1}{d}D'(1 - \frac{K'(x)}{k})$. Then $g'(x) = \frac{1}{dk}D''(1 - \frac{K'(x)}{k})K''(x)$ which is nonnegative over $[0, 1]$. We also have $\phi(x) = g(x) - x$, where ϕ is given in (1.4). Since $\phi(1) = -D'(0)/d \leq 0$, $\phi'(\rho) < 0$ by the hypothesis, and $g(x)$ is nondecreasing in $[0, 1]$, it follows that $|g'(\rho)| < 1$ and thus ρ is an attractive fix point of $x = g(x)$. As $\rho_0 = 1$. It follows that $\rho_t \to \rho$ as $t \to \infty$. Consequently, for every $\hat{\varepsilon} > 0$ there is sufficiently large I such that $|\rho_t - \rho| < \hat{\varepsilon}$. Hence, after I iterations of the parallel stripping process, the number of vertices remaining is $(\lambda + o(1))n + O_n(\hat{\varepsilon}n)$ where

$$\lambda = 1 - D\left(1 - \frac{K'(\rho)}{k}\right) - \frac{K'(\rho)}{k} D'\left(1 - \frac{K'(\rho)}{k}\right). \tag{6.3}$$

If $\rho = 0$ then $\lambda = 0$ by Remark 1.3(c). Our theorem for $\boldsymbol{n}^*$ follows by letting $I \to \infty$. Since $\boldsymbol{k} \geq 1$, $K(0) = 0$ and thus $\boldsymbol{m}^*/n = \frac{d}{k}K(0) + o_{\hat{\varepsilon},n}(1) = o_{\hat{\varepsilon},n}(1)$. This establishes (1.6) when $\rho = 0$.

Suppose $\rho > 0$. It is sufficient to show that the 2-core is obtained after further removing $O_n(\hat{\varepsilon}n)$ vertices, following the same approach as [47, lemma 4]. We briefly sketch it. Following the same argument as before, the probability that a random vertex has degree $j \geq 2$ after I iterations of the stripping process is

$$\sum_{i \geq j} \mathbb{P}(\boldsymbol{d} = i)\binom{i}{j}(\mathbb{E}\rho_I^{\hat{\boldsymbol{k}}-1})^j(1 - \mathbb{E}\rho_I^{\hat{\boldsymbol{k}}-1})^{i-j} = \sum_{i \geq j} \mathbb{P}(\boldsymbol{d} = i)\binom{i}{j}\left(\frac{K'(\rho_I)}{k}\right)^j\left(1 - \frac{K'(\rho_I)}{k}\right)^{i-j}.$$

Similarly, the probability of a uniformly random hyperedge in G having size $j \geq 1$ and surviving the first I iterations of the stripping process is

$$\mathbb{P}(\boldsymbol{k} = j)\rho_I^j.$$

The number of vertices with degree less than 2 after I iterations is bounded by $(\lambda_I - \lambda_{I+1})n + o_n(n)$. Hence, by choosing I sufficiently large, we can make these quantities arbitrarily close to those with ρ_I replaced by ρ. Now standard concentration arguments apply to show that the number of degree $j \geq 2$ vertices is $\gamma_j n + O_n(\hat{\varepsilon}n)$, where

$$\gamma_j = \sum_{i=0}^{\infty} \mathbb{P}(\boldsymbol{d} = i)\binom{i}{j}\left(\frac{K'(\rho)}{k}\right)^j\left(1 - \frac{K'(\rho)}{k}\right)^{i-j},$$

the number of vertices of degree less than 2 is $O_n(\hat{\varepsilon}n)$. Similarly, the number of remaining hyperedges of size j is $\mathbb{P}(\boldsymbol{k} = j)\rho^j m + O_n(\hat{\varepsilon}n)$, and the total degree of the remaining vertices is

$$m\sum_{j \geq 1} j\mathbb{P}(\boldsymbol{k} = j)\rho^j + O_n(\hat{\varepsilon}n) = m\rho K'(\rho) + O_n(\hat{\varepsilon}n) = (dn/k)\rho K'(\rho) + O_n(\hat{\varepsilon}n). \tag{6.4}$$

Note that $\hat{\varepsilon}$ can be made arbitrarily small by choosing sufficiently large I.

Now we remove one hyperedge incident with a vertex with degree 1 at a time. Call this process SLOWSTRIP. Let G_t denote the hypergraph obtained after t steps of SLOWSTRIP and let X_t denote the total degree of the vertices of degree 1 in G_t, Then, for all $t = O_n(\hat{\varepsilon}n)$ such that $X_t > 0$:

$$\begin{aligned}
&\mathbb{E}(X_{t+1} - X_t | G_t)\\
&= -1 + \sum_{j\geq 1} \frac{j\mathbb{P}(\boldsymbol{k} = j)\rho^j m}{\rho K'(\rho)m} \cdot (j-1) \cdot \frac{2\gamma_2 n}{(dn/k)\rho K'(\rho)} + O_n(\hat{\varepsilon})\\
&= -1 + \frac{1}{\rho K'(\rho)} \left(\sum_{j\geq 1} j(j-1)\mathbb{P}(\boldsymbol{k} = j)\rho^j \right) \frac{2 \cdot \frac{1}{2}(K'(\rho)/k)^2 D''(1 - K'(\rho)/k)}{K'(\rho) d\rho/k} + O_n(\hat{\varepsilon})\\
&= -1 + \frac{D''(1 - K'(\rho)/k)K''(\rho)}{kd} + O_n(\hat{\varepsilon}).
\end{aligned}$$

Note that in the first equation above, -1 accounts for the removal of one variable clone x from the set of vertices of degree less than 2. The term $j\mathbb{P}(\boldsymbol{k} = j)\rho^j m/\rho K'(\rho)m$ approximates the probability that x is contained in a hyperedge of size j, up to an $O_n(\hat{\varepsilon})$ error. In that case, $j-1$ variable clones that lie in the same hyperedge as x will be removed. For each of these $j-1$ deleted variable clones, if it lies in a variable of degree 2, then it results in one new variable node of degree 1. The probability for that to happen is approximated by $2\gamma_2 n/D_t$, up to an $O_n(\hat{\varepsilon})$ error, where D_t denotes the total degree of G_t and by (6.4), $D_t = (dn/k)\rho K'(\rho) + O_n(\hat{\varepsilon}n)$. For the second equation above, note that

$$\gamma_2 = \sum_{i\geq 2} \mathbb{P}(\boldsymbol{d} = i) \frac{i(i-1)}{2} \left(\frac{K'(\rho_l)}{k}\right)^2 \left(1 - \frac{K'(\rho_l)}{k}\right)^{i-2} + O_n(\hat{\varepsilon}) = \frac{1}{2}\left(\frac{K'(\rho)}{k}\right)^2 D''(1 - K'(\rho)/k) + O_n(\hat{\varepsilon}).$$

By the assumption that $\phi'(\rho) < 0$ we have

$$-1 + \frac{D''(1 - K'(\rho)/k)K''(\rho)}{kd} < 0.$$

Hence, $\mathbb{E}(X_{t+1} - X_t | G_t) < -\delta$ for some $\delta > 0$, by making $\hat{\varepsilon}$ sufficiently small (i.e. by choosing sufficiently large I). Then the standard Azuma inequality [lemma 29] (with Lipschitz constant $(n \log n)^{1/(2+\sigma)}$ by Claim 6.1 will be sufficient to show that X_t decreases to 0 after $O_n(\hat{\varepsilon}n) = o_{\hat{\varepsilon},n}(n)$ steps (See details in [47, lemma 4]). The case $\rho > 0$ of the theorem follows by

$$\lim_{n\to\infty} \frac{\boldsymbol{m}^*}{n} = \lim_{n\to\infty} \frac{m}{n} \cdot \sum_{j\geq 1} \mathbb{P}(\boldsymbol{k} = j)\rho^j = \frac{d}{k} K(\rho),$$

as desired. This proves (1.6) when $\boldsymbol{k} \geq 1$.

Suppose now that $p_0 = \mathbb{P}(\boldsymbol{k} = 0) > 0$. Let $\overline{G}$ be the hypergraph obtained from G by deleting all hyperedges with size 0. Let $\overline{m}$ denote the number of hyperedges in $\overline{G}$. Then, with probability $1 - o_n(1)$, $\overline{m} \sim (1 - p_0)m$. The size of a uniformly random hyperedge in $\overline{G}$ has the same distribution as $\overline{\boldsymbol{k}}$, defined by $\boldsymbol{k}$ conditioned on $\boldsymbol{k} \geq 1$. Let $\overline{k} = \mathbb{E}\overline{\boldsymbol{k}}$. Then, $\overline{k} = k/(1 - p_0)$. Let $\overline{K}(\alpha)$ be the probability generating function of $\overline{\boldsymbol{k}}$. Then, immediately

$$\overline{K}(\alpha) = \frac{K(\alpha) - p_0}{1 - p_0}, \quad \overline{K}'(\alpha) = \frac{K'(\alpha)}{1 - p_0}, \quad \overline{K}''(\alpha) = \frac{K''(\alpha)}{1 - p_0}.$$

Let $\overline{\Phi}$ be the function obtained from Φ by replacing $K(\alpha)$, $K'(\alpha)$ and k by $\overline{K}(\alpha)$, $\overline{K}'(\alpha)$ and $\overline{k}$, respectively. It is straightforward to see that the set of stable points of Φ corresponds to the set of stable points of $\overline{\Phi}$. Thus, by letting $\overline{\rho} = \max\{x \in [0,1] : \overline{\Phi}'(x) = 1\}$ it follows then that $\overline{\rho} = \rho$. Applying (1.6) with $\overline{\boldsymbol{k}} \geq 1$ to $\overline{G}$,

$$\begin{aligned}\lim_{n\to\infty} \frac{\overline{\boldsymbol{n}}^*}{n} &= 1 - D\left(1 - \frac{\overline{K}'(\overline{\rho})}{\overline{k}}\right) - \frac{\overline{K}'(\overline{\rho})}{\overline{k}} D'\left(1 - \frac{\overline{K}'(\overline{\rho})}{\overline{k}}\right)\\ &= 1 - D\left(1 - \frac{K'(\rho)}{k}\right) - \frac{K'(\rho)}{k} D'\left(1 - \frac{K'(\rho)}{k}\right)\\ \lim_{n\to\infty} \frac{\overline{\boldsymbol{m}}^*}{n} &= \frac{d}{\overline{k}}\overline{K}(\overline{\rho}) = \frac{d}{k}(K(\rho) - p_0),\end{aligned}$$

where $\overline{\boldsymbol{n}}^*$ and $\overline{\boldsymbol{m}}^*$ denote the numbers of vertices and hyperedges in $\overline{G}$. Since $\boldsymbol{n}^* = \overline{\boldsymbol{n}}^*$ and $\boldsymbol{m}^* = \overline{\boldsymbol{m}}^* + (1 + o_n(1))p_0 m = (1 + o_n(1))(\overline{\boldsymbol{m}}^* + p_0 dn/k)$, as the set of hyperedges of size 0 in G remain in the 2-core of G, the equations (1.6) holds as well for the case that $p_0 > 0$. □

Proof of Claim 6.1. Since both $\mathbb{E}\boldsymbol{d}^{2+\sigma} = O_n(1)$ and $\mathbb{E}\boldsymbol{k}^{2+\sigma} = O_n(1)$, the probability that $\boldsymbol{d} > (n\log n)^{1/(2+\sigma)}$ or $\boldsymbol{k} > (n\log n)^{1/(2+\sigma)}$ is $O_n(1/n\log n)$. The bound on the maximum degree and maximum size of the hyperedges in $\boldsymbol{G}$ follows by taking the union bound.

For any $u \in [n]$, let $N_i(u) = |\partial^i(u)|$. We will prove that with high probability for every u and for every fixed i, $N_i(u) = O_n(n^{1/(2+\sigma)}\log^2 n)$, which then completes the proof for Claim 6.1. We prove by induction. Let $d_1, \ldots, d_{N_i(u)}$ denote the degrees of the vertices in $\partial^i(u)$. Then the number of hyperedges incident with these vertices is bounded by $M := \sum_{j=1}^{N_i(u)} d_j$. By the construction of $\boldsymbol{G}$, each M is stochastically dominated by $\sum_{j=1}^{N_i(u)}(1 + o_n(1))\hat{\boldsymbol{d}}_j$ where $\hat{\boldsymbol{d}}_j$ are i.i.d. copies of $\hat{\boldsymbol{d}}$ whose distribution is given in (6.2). The $o_n(1)$ error is caused by the exposure of $\cup_{j\leq i}\partial^j(u)$ which contains $o_n(n)$ vertices by induction. Since $\mathbb{E}\boldsymbol{d}^{2+\sigma} = O_n(1)$, we have $\hat{d} := \mathbb{E}\hat{\boldsymbol{d}} = O_n(1)$. Note that $\mathbb{E}M = \hat{d}N_i(u)$. Applying the Chernoff bound to the sum of independent $[0,1]$-valued random variables we have

$$\begin{aligned}&\mathbb{P}\left(M \geq 2\hat{d}N_i(u) + n^{1/(2+\sigma)}\log^2 n\right)\\ &\quad= \mathbb{P}\left(\sum_{j=1}^{N_i(u)} \frac{\hat{\boldsymbol{d}}_i}{(n\log n)^{1/(2+\sigma)}} \geq \frac{2\hat{d}N_i(u)}{(n\log n)^{1/(2+\sigma)}} + (\log n)^{(3+\sigma)/(2+\sigma)}\right) < n^{-2}.\end{aligned}$$

Similarly, $N_{i+1}(u)$ is bounded by $\sum_{j=1}^{M} k_i$, where k_i are the sizes of the hyperedges incident with the vertices in $\partial^i(u)$. Similarly, $\sum_{j=1}^{M} k_i$ is stochastically dominated by $(1+o_n(1))\sum_{j=1}^{M}\hat{\boldsymbol{k}}_j$ where $\hat{\boldsymbol{k}}_j$ are i.i.d. copies of $\hat{\boldsymbol{k}}$ whose distribution is defined in (6.1). Let $\hat{k} = \mathbb{E}\hat{\boldsymbol{k}}$. Applying the Chernoff bound again we obtain that with probability at least $1-n^{-2}$, $N_{i+1}(u) < 2\hat{k}M + n^{1/(2+\sigma)}\log^2 n < 4\hat{d}\hat{k}N_i(u) + (1+2\hat{k})n^{1/(2+\sigma)}\log^2 n$. Apply this recursion inductively and the union bound on the failure probability, we obtain $N_i(u) = O_n(n^{1/(2+\sigma)}\log^2 n)$, as desired. □

Proof of Claim 6.2. Fix $\varepsilon > 0$. Choose $L = L(\varepsilon, r)$ sufficiently large so that the probability that $d_v > L$ is smaller than ε (note that v is a uniformly random vertex). Given $d_v \leq L$. Let $k_1, \ldots, k_{d_v}$ be the sizes of the hyperedges incident to v. Similarly to the proof of Claim 6.1, k_js are approximated by i.i.d. copies of $\hat{\boldsymbol{k}}$ defined in (6.1), up to an $1+o(1)$ multiplicative error. We can assume L is sufficiently large so that with probability at least $1-\varepsilon$, $\sum_{i=1}^{d_v} k_i \leq L$. Inductively, we can make L sufficiently large so that $|\partial^i(v)| \leq L$ for all $i \leq R$. Let $\mathscr{E}_i$ denote the set of hyperedges incident with vertices in $\partial^i(v)$, but not incident with any in $\partial^{i-1}(v)$. Cycles in $\partial^i(v)$ can appear in two ways: (a) two vertices in $\partial^i(v)$ are incident with the same hyperedge in $\mathscr{E}_i$; (b) two hyperedges in $\mathscr{E}_{i-1}$ are incident with the same vertex

in $\partial^i(v)$. We will prove that with high probability, none of the two cases occurs for any fixed i. For (a), let $(d_j)_{j\in\partial^i(v)}$ denote the degrees of the vertices in $\partial^i(v)$. The expected number of occurrences of pairs of vertices in (a) is

$$\mathbb{E}\left(\sum_{j,k\in\partial^i(v)}\binom{d_j}{2}\binom{d_k}{2}\sum_{h\in[\boldsymbol{m}]}\binom{k_h}{2}O_n(n^{-2})\right)=O_n(n^{-1})\mathbb{E}\left(\sum_{j,k\in\partial^i(v)}d_j^2d_k^2\right). \tag{6.5}$$

Note that $|\partial^i(v)|\le L$ for each $j\le R$. This immediately implies that $d_j\le L$ for all $j\in\partial^i(v)$. Hence, the above probability is $O_n(n^{-1})$. The probability that $|\partial^i(v)|\le L$ fails is at most $R\varepsilon$ by our choice of L. Hence, the probability that (a) fails is at most $R\varepsilon+o_n(1)$. The treatment of (b) is analogous. Our claim now follows by letting $\varepsilon\to 0$. □

7 | PROOF OF THEOREM 1.4

Recall that

$$\phi(\alpha)=1-\alpha-\frac{1}{d}D'\left(1-\frac{K'(\alpha)}{k}\right). \tag{7.1}$$

For Theorem 1.4 and Remark 1.5, it is sufficient to prove that if condition (i) or (ii) is satisfied then (a) $\max_{\alpha\in[0,1]}\Phi(\alpha)=\max\{\Phi(0),\Phi(\rho)\}$; and (b) $\phi'(\rho)<0$ unless

$$\mathbb{P}(\boldsymbol{d}=1)=0\quad\text{and}\quad 2(\mathbb{E}\boldsymbol{k}-1)\mathbb{P}(\boldsymbol{d}=2)>\mathbb{E}\boldsymbol{d}. \tag{7.2}$$

Since $\Phi(\alpha)$ is continuous on $[0,1]$, the maximum occurs at either 0 or 1 or at a stable point.

Case A: $\mathrm{Var}(\boldsymbol{k})=0$. In this case, $\boldsymbol{k}=k$ always and thus $K(\alpha)=\alpha^k$. We must have $k\ge 1$ since otherwise $k=d=0$. If $k=1$ then $\phi'(\alpha)=-1$ which implies (b) immediately. Moreover, $K''(\alpha)=0$ for all $\alpha\in[0,1]$ and thus $\Phi'(\alpha)=(d/k)K''(\alpha)\phi(\alpha)=0$ for all $\alpha\in[0,1]$. This implies (a).

Next consider the case that $k=2$. Then, $\phi''(\alpha)=-\frac{1}{d}D'''(1-\alpha)<0$ on $(0,1)$ unless $\boldsymbol{d}\le 2$. Consider the case that $\operatorname{supp}\boldsymbol{d}\cap\mathbb{N}_{\ge 3}\neq\emptyset$. Then ϕ is concave and can have at most 2 roots. Obviously $\alpha=0$ is a root. Let ρ denote the other root if exists. We must have $\phi'(\rho)<0$ by the concavity of ϕ. Hence, the maximum of Φ cannot be achieved at 1. Thus, the maximas of Φ can only be from $\{0,\rho\}$. This verifies (a) and (b). Now assume $\boldsymbol{d}\le 2$. Then $\phi''(\alpha)=0$ on $[0,1]$. Hence $\phi'(\alpha)=\phi'(1)=-1+\mathbb{P}(\boldsymbol{d}=2)/d<0$ for all $\alpha\in[0,1]$. Thus, $\phi(\alpha)$ is a line with a negative slope and has exactly one root at 0 on $[0,1]$. Hence $\rho=0$ and $\phi'(\rho)<0$. This verifies (a) and (b).

Next we consider the case that $k\ge 3$. We have

$$\begin{aligned}
\phi(\alpha)&=1-\alpha-\frac{1}{d}D'(1-\alpha^{k-1})\\
\phi'(\alpha)&=-1+\frac{(k-1)\alpha^{k-2}}{d}D''(1-\alpha^{k-1})\\
\phi''(\alpha)&=\frac{k-1}{d}\alpha^{k-3}\left((k-2)D''(1-\alpha^{k-1})-(k-1)D'''(1-\alpha^{k-1})\alpha^{k-1}\right)\\
&=\frac{k-1}{d}\alpha^{k-3}\left((k-2)D''(t)-(k-1)D'''(t)(1-t)\right)\quad\text{where } t=1-\alpha^{k-1}.
\end{aligned}$$

Hence,

$$\phi(0)=0\qquad\phi(1)=-\frac{1}{d}D'(0)\le 0 \tag{7.3}$$

$$\phi'(0)=-1\qquad\phi'(1)=-1+\frac{k-1}{d}D''(0). \tag{7.4}$$

Recall that $\Phi'(\alpha) = \frac{d}{k}K''(\alpha)\phi(\alpha)$. We have $K''(\alpha) > 0$ for all $\alpha \in (0,1]$. By (7.3) we have $\Phi'(1) \le 0$ and thus the supremum of $\Phi(\alpha)$ can only occur at 0 or a stable point. In all of the following subcases, we will prove that $\phi''(\alpha)$ has at most 1 root in $[0,1]$ (except for some trivial cases that we discuss separately). It follows immediately that ϕ can have at most three roots on $[0,1]$ including the trivial one at $\alpha = 0$. Now we prove that this implies claims (a) and (b).

If ϕ has only a trivial root, then so is $\Phi'(\alpha)$. Thus, $\alpha = 0$ is the unique maxima of $\Phi(\alpha)$ and $\rho = 0$. This verifies (a). As $\phi'(0) = -1$ we immediately have $\phi'(\rho) < 0$.

If ϕ has two roots, then the larger root is ρ. Since $\phi'(0) < 0$, in this case, ϕ is negative in $(0,\rho)$ and positive in $(\rho,1)$. This is only possible when $\phi(1) = 0$ by (7.3), which requires $\mathbb{P}(\boldsymbol{d} = 1) = 0$. In this case, $\rho = 1$. Next we consider two further cases: (i) $2(k-1)\mathbb{P}(\boldsymbol{d} = 2) > d$ corresponding to $\phi'(1) > 0$; (ii) $2(k-1)\mathbb{P}(\boldsymbol{d} = 2) < d$ corresponding to $\phi'(1) < 0$. As ϕ has only two roots, case (ii) obviously cannot happen. Thus, it means that the only situation that ϕ has two roots would be $\mathbb{P}(\boldsymbol{d} = 1) = 0$ and $2(k-1)\mathbb{P}(\boldsymbol{d} = 2) > d$, as in (7.2). In this situation we are only required to verify (a). Note that ϕ is negative in $(0,1)$ as $\rho = 1$. It follows then that $\Phi(\alpha)$ is a decreasing function in $(0,1)$. Hence, $\alpha = 0$ is the unique maxima, as desired.

If ϕ has three roots, then there is a root ρ^* between 0 and ρ. Then ϕ is negative in $(0,\rho^*)$ and positive in (ρ^*,ρ). As $K''(\alpha) > 0$ for all $\alpha \in (0,1]$, the sign of ϕ implies that ρ^* is a local minima and ρ is a local maxima. This verifies (a). Moreover, as ϕ is positive in (ρ^*,ρ) and $\phi(\rho) = 0$, $\phi'(\rho) < 0$ follows immediately.

Case A1: $\mathrm{Var}(\boldsymbol{k}) = 0$ and $\mathrm{Var}(\boldsymbol{d}) = 0$. In this case $\boldsymbol{d} = d$. Then $D(\alpha) = \alpha^d$. If $d \ge 3$ then

$$\begin{aligned}\phi''(\alpha) &= \frac{k-1}{d}\alpha^{k-3}\left((k-2)d(d-1)t^{d-2} - (k-1)d(d-1)(d-2)t^{d-3}(1-t)\right)\\ &= (k-1)(d-1)t^{d-3}\alpha^{k-3}\left((k-2)t - (k-1)(d-2)(1-t)\right) \quad \text{where } t = 1-\alpha^{k-1}.\end{aligned}$$

Obviously, $\phi''(\alpha)$ has a unique root in $[0,1]$.

If $d = 1$ then $\phi'(\alpha) = -1$ and so ϕ has only a trivial root at $\alpha = 0$; If $d = 2$ then $\phi''(\alpha) > 0$ in $(0,1)$ and so ϕ is convex and thus has only a trivial root at $\alpha = 0$ by (7.3). Hence for $d \le 2$, $\rho = 0$ and is the unique maxima. Claims (a) and (b) hold trivially.

Case A2: $\mathrm{Var}(\boldsymbol{k}) = 0$ and $\boldsymbol{d} \sim \mathbf{Po}_{\ge r}(\lambda)$. In this case $D(\alpha) = h_r(\lambda\alpha)/h_r(\lambda)$, where

$$h_r(x) = \sum_{j \ge r}\frac{x^j}{j!} \quad \text{for all nonnegative integers } r; \tag{7.5}$$

$$h_r(x) = e^x \quad \text{for all negative integers } r. \tag{7.6}$$

Then, for all integers t,

$$D'(\alpha) = \frac{\lambda h_{r-1}(\lambda\alpha)}{h_r(\lambda)},\ D''(\alpha) = \frac{\lambda^2 h_{r-2}(\lambda\alpha)}{h_r(\lambda)},\ D'''(\alpha) = \frac{\lambda^3 h_{r-3}(\lambda\alpha)}{h_r(\lambda)}.$$

Since $\mathbb{E}\boldsymbol{d} = d$, it requires that λ satisfies

$$D'(1) = \frac{\lambda h_{r-1}(\lambda)}{h_r(\lambda)} = d. \tag{7.7}$$

Thus,

$$\phi''(\alpha) = \frac{(k-1)d\alpha^{k-3}}{h_r(\lambda)}\left((k-2)h_{r-2}(\lambda t) - (k-1)(1-t)h_{r-3}(\lambda t)\right).$$

Solving $\phi''(\alpha) = 0$ yields

$$\frac{k-1}{k-2}(1-t) = \frac{h_{r-2}(\lambda t)}{h_{r-3}(\lambda t)} = 1 - \frac{h_{r-3}(\lambda t) - h_{r-2}(\lambda t)}{h_{r-3}(\lambda t)}. \tag{7.8}$$

The right-hand side above is obviously a constant function if $r \le 2$. If $r \ge 3$, then $h_{r-3}(\lambda t) - h_{r-2}(\lambda t) = (\lambda t)^{r-3}/(r-3)!$, and $h_{r-3}(\lambda t)$ is a power series of λt with minimum degree $r-3$. Hence, by dividing $(\lambda t)^{r-3}/(r-3)!$ from both the numerator and the denominator, we immediately get that the right-hand side of (7.8) is an increasing function. However the left-hand side of (7.8) is a decreasing function. Hence (7.8) has at most one solution, implying that $\phi''(\alpha)$ has at most one root.

Case B: $\boldsymbol{k} \sim \mathbf{Po}_{\ge s}(\gamma)$. We must have γ satisfy

$$\frac{\gamma h_{s-1}(\gamma)}{h_s(\gamma)} = k,$$

so that $\mathbb{E}\boldsymbol{k} = k$. Here $k > s$ is required (to guarantee the existence of γ if $s \ge 1$, and to avoid triviality if $s = 0$). Now we have $K(\alpha) = h_s(\gamma\alpha)/h_s(\gamma)$, where h_s is defined as in (7.5) and (7.6). Thus,

$$\begin{aligned}
\phi(\alpha) &= 1 - \alpha - \frac{1}{d}D'\left(1 - \frac{h_{s-1}(\gamma\alpha)}{h_{s-1}(\gamma)}\right)\\
\phi'(\alpha) &= -1 + \frac{\gamma h_{s-2}(\gamma\alpha)}{d h_{s-1}(\gamma)}D''\left(1 - \frac{h_{s-1}(\gamma\alpha)}{h_{s-1}(\gamma)}\right)\\
\phi''(\alpha) &= \frac{\gamma^2}{d h_{s-1}(\gamma)}\left(h_{s-3}(\gamma\alpha)D''\left(1 - \frac{h_{s-1}(\gamma\alpha)}{h_{s-1}(\gamma)}\right) - \frac{h_{s-2}(\gamma\alpha)^2}{h_{s-1}(\gamma)}D'''\left(1 - \frac{h_{s-1}(\gamma\alpha)}{h_{s-1}(\gamma)}\right)\right).
\end{aligned}$$

Hence,

$$\phi(0) = 0 \qquad \phi(1) = -\frac{1}{d}D'(0) \le 0 \tag{7.9}$$

$$\phi'(0) = -1 \qquad \phi'(1) = -1 + \frac{\gamma h_{s-2}(\gamma)}{d h_{s-1}(\gamma)}D''(0). \tag{7.10}$$

As before, we will prove that $\phi''(\alpha)$ has at most 1 root in $[0,1]$ (except for some trivial cases that will be discussed separately), which is sufficient to ensure (a) and (b).

Case B1: $\boldsymbol{k} \sim \mathbf{Po}_{\ge s}(\gamma)$ and $\mathrm{Var}(\boldsymbol{d}) = 0$. In this case $\boldsymbol{d} = d$. Then $D(\alpha) = \alpha^d$. If $d \ge 3$ then solving $\phi''(\alpha) = 0$ yields

$$\frac{d-2}{h_{s-1}(\gamma)} \cdot h_{s-2}(\gamma\alpha) = \left(1 - \frac{h_{s-1}(\gamma\alpha)}{h_{s-1}(\gamma)}\right)\frac{h_{s-3}(\gamma\alpha)}{h_{s-2}(\gamma\alpha)}. \tag{7.11}$$

On the right hand side above, $1 - h_{s-1}(\gamma\alpha)/h_{s-1}(\gamma) \ge 0$ and is a decreasing function of α. We also have

$$\frac{h_{s-3}(\gamma\alpha)}{h_{s-2}(\gamma\alpha)} = \frac{h_{s-3}(\gamma\alpha)}{h_{s-3}(\gamma\alpha) - (\gamma\alpha)^{s-3}/(s-3)!} = \left(1 - \frac{(\gamma\alpha)^{s-3}/(s-3)!}{h_{s-3}(\gamma\alpha)}\right)^{-1},$$

which is positive and a decreasing function of α if $s \ge 3$, and is equal to 1 if $s \le 2$. Hence, the left-hand side of (7.11) is an increasing function whereas the right hand side is a decreasing function. Hence $\phi''(\alpha)$ has at most one root.

If $d \le 2$ the same argument as in Case A1 shows that claims (a) and (b) hold.

Case B2: $\boldsymbol{k} \sim \mathbf{Po}_{\ge s}(\gamma)$ and $\boldsymbol{d} \sim \mathbf{Po}_{\ge r}(\lambda)$. In this case $D(\alpha) = h_r(\lambda\alpha)/h_r(\lambda)$, and λ necessarily satisfies (7.7). Then solving $\phi''(\alpha) = 0$ yields

$$\frac{\lambda}{h_{s-1}(\gamma)}h_{s-2}(\gamma\alpha) = \frac{h_{s-3}(\gamma\alpha)}{h_{s-2}(\gamma\alpha)} \cdot \frac{h_{r-2}(\lambda(1 - h_{s-1}(\gamma\alpha)/h_{s-1}(\gamma)))}{h_{r-3}(\lambda(1 - h_{s-1}(\gamma\alpha)/h_{s-1}(\gamma)))}$$

The left-hand side is an increasing function whereas the right hand side is the product of two functions, both of which are either equal to 1 or a positive decreasing function. Thus, $\phi''(\alpha)$ has at most one root.

ACKNOWLEDGMENT

We thank David Saad for a helpful conversation and Guilhem Semerjian for bringing [41] to our attention. Open access funding enabled and organized by Projekt DEAL.

REFERENCES

1. D. Achlioptas and M. Molloy, *The solution space geometry of random linear equations*, Random Struct. Algorithms **46** (2015), 197–231.
2. M. Aizenman, R. Sims, and S. Starr, *An extended variational principle for the SK spin-glass model*, Phys. Rev. B **68** (2003), 214403.
3. R. Alamino and D. Saad, *Typical kernel size and number of sparse random matrices over Galois fields: A statistical physics approach*, Phys. Rev. E **77** (2008), 061123.
4. D. Aldous, *Representations for partially exchangeable arrays of random variables*, J. Multivar. Anal. **11** (1981), 581–598.
5. D. Aldous and M. Steele, "*The objective method: Probabilistic combinatorial optimization and local weak convergence*," Probability on discrete structures, H. Kesten (ed.), Springer, New York, NY, 2003, 2004.
6. N. Alon and J. Spencer, The probabilistic method, 2nd ed., Wiley, Hoboken, NJ, 2000.
7. P. Ayre, A. Coja-Oghlan, P. Gao, and N. Müller, *The satisfiability threshold for random linear equations*, Combinatorica **40** (2020), 179–235.
8. G. Balakin, *On random matrices*, Theory Probab. Appl. **12** (1967), 346–353.
9. G. Balakin, *The distribution of random matrices over a finite field*, Theory Probab. Appl. **13** (1968), 631–641.
10. V. Bapst and A. Coja-Oghlan, *Harnessing the Bethe free energy*, Random Struct. Algorithms **49** (2016), 694–741.
11. M. Bayati, D. Gamarnik, and P. Tetali, *Combinatorial approach to the interpolation method and scaling limits in sparse random graphs*, Ann. Probab. **41** (2013), 4080–4115.
12. J. Blömer, R. Karp, and E. Welzl, *The rank of sparse random matrices over finite fields*, Random Struct. Algorithms **10** (1997), 407–419.
13. C. Bordenave, M. Lelarge, and J. Salez, *The rank of diluted random graphs*, Ann. Probab. **39** (2011), 1097–1121.
14. C. Bordenave, M. Lelarge, and J. Salez, *Matchings on infinite graphs*, Probab. Theory Related Fields **157** (2013), 183–208.
15. C. Bordenave, *A new proof of Friedman's second eigenvalue theorem and its extension to random lifts*, Ann. Sci. Ecole. Norm. S. **53** (2020), 1393–1439.
16. A. Coja-Oghlan, O. Cooley, M. Kang, J. Lee, and J. Ravelomanana, The sparse parity matrix, 2021. arXiv:2107.06123.
17. A. Coja-Oghlan, F. Krzakala, W. Perkins, and L. Zdeborova, *Information-theoretic thresholds from the cavity method*, Adv. Math. **333** (2018), 694–795.
18. A. Coja-Oghlan and W. Perkins, *Spin systems on Bethe lattices*, Commun. Math. Phys. **372** (2019), 441–523.
19. C. Cooper, *The cores of random hypergraphs with a given degree sequence*, Random Struct. Algorithms **25** (2004), no. 4, 353–375.
20. C. Cooper, A. Frieze, and W. Pegden, *On the rank of a random binary matrix*, Electron. J. Comb. **26** (2019), P4.12.
21. K. Costello and V. Vu, *The rank of random graphs*, Random Struct. Algorithms **33** (2008), 269–285.
22. K. Costello and V. Vu, *On the rank of random sparse matrices*, Comb. Probab. Comput. **19** (2010), 321–342.
23. M. Dietzfelbinger, A. Goerdt, M. Mitzenmacher, A. Montanari, R. Pagh, and M. Rink, *Tight thresholds for cuckoo hashing via XORSAT,* Proceedings of the 37th ICALP, 2010. pp. 213–225.
24. O. Dubois and J. Mandler, *The 3-XORSAT threshold,* Proc. 43rd FOCS, 2002, pp. 769–778.
25. S. Franz and M. Leone, *Replica bounds for optimization problems and diluted spin systems*, J. Stat. Phys. **111** (2003), 535–564.
26. J. Friedman, *A proof of Alon's second eigenvalue conjecture and related problems*, Mem. Am. Math. Soc. **195** (2008), 1–114.
27. J. Fulman and L. Goldstein, *Stein's method and the rank distribution of random matrices over finite fields*, Ann. Probab. **43** (2015), 1274–1314.
28. A. Giurgiu, N. Macris, and R. Urbanke, *Spatial coupling as a proof technique and three applications*, IEEE Trans. Inf. Theory **62** (2016), 5281–5295.
29. A. Goerdt and L. Falke, *Satisfiability thresholds beyond k-XORSAT*, Proc. 7th Int. Comput. Sci. Sympos. Russia, 2012, pp. 148–159.

30. F. Guerra, *Broken replica symmetry bounds in the mean field spin glass model*, Commun. Math. Phys. **233** (2003), 1–12.
31. D. Hoover, Relations on probability spaces and arrays of random variables. Preprint, Institute of Advanced Studies, Princeton, NJ, 1979.
32. M. Ibrahimi, Y. Kanoria, M. Kraning, and A. Montanari, *The set of solutions of random XORSAT formulae*, Ann. Appl. Probab. **25** (2015), 2743–2808.
33. V. Jain, A. Sah, and M. Sawhney, *Singularity of discrete random matrices*, Geometr. Funct. Anal. **31** (2021), 1160–1218.
34. J. Kahn, J. Komloś, and E. Szemerédi, *On the probability that a random ±1-matrix is singular*, J. AMS **8** (1995), 223–240.
35. O. Kallenberg, Probabilistic symmetries and invariance principles, Springer, New York, NY, 2005.
36. J. Komlós, *On the determinant of (0,1) matrices*, Studia Sci. Math. Hungar. **2** (1967), 7–21.
37. I. Kovalenko, *On the limit distribution of the number of solutions of a random system of linear equations in the class of Boolean functions*, Theory Probab. Appl. **12** (1967), 51–61.
38. I. Kovalenko, A. Levitskaya, and M. Savchuk, Selected problems of probabilistic combinatorics, Naukova Dumka, Kiev, 1986.
39. F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, and P. Zhang, *Spectral redemption in clustering sparse networks*, Proc. Natl. Acad. Sci. **110** (2013), 20935–20940.
40. S. Kumar, A. Young, N. Macris, and H. Pfister, *Threshold saturation for spatially-coupled LDPC and LDGM codes on BMS channels*, IEEE Trans. Inf. Theory **60** (2014), 7389–7415.
41. M. Lelarge, *Bypassing correlation decay for matchings with an application to XORSAT*, Proc. IEEE Inf. Theory Workshop, 2013, pp. 1–5.
42. A. Levitskaya, *Theorems on invariance for the systems of random linear equations over an arbitrary finite ring*, Soviet Math. Dokl. **263** (1982), 289–291.
43. A. Levitskaya, *The probability of consistency of a system of random linear equations over a finite ring*, Theory Probab. Appl. **30** (1985), 339–350.
44. A. Litvak and K. Tikhomirov, Singularity of sparse Bernoulli matrices, 2020. arXiv: 2004.03131.
45. M. Mehta, Random matrices, Elsevier Academic Press, Amsterdam, 2004.
46. M. Mézard and A. Montanari, Information, physics and computation, Oxford University Press, Oxford, UK, 2009.
47. M. Molloy, *Cores in random hypergraphs and Boolean formulas*, Random Struct. Algorithms **27** (2005), 124–135.
48. A. Montanari, *Estimating random variables from random sparse observations*, Eur. Trans. Telecommun. **19** (2008), no. 4, 385–403.
49. D. Panchenko, *Spin glass models from the point of view of spin distributions*, Ann. Probab. **41** (2013), 1315–1361.
50. B. Pittel and G. Sorkin, *The satisfiability threshold for k-XORSAT*, Comb. Probab. Comput. **25** (2016), 236–268.
51. P. Raghavendra and N. Tan, *Approximating CSPs with global cardinality constraints using SDP hierarchies,* Proc. 23rd SODA, 2012, pp. 373–387.
52. T. Richardson and R. Urbanke, Modern coding theory, Cambridge University Press, Cambridge, UK, 2008.
53. T. Tao and V. Vu, *On the singularity probability of random Bernoulli matrices*, J. AMS **20** (2007), 603–628.
54. K. Tikhomirov, *Singularity of random Bernoulli matrices*, Ann. Math. **191** (2020), 593–634.
55. H. van Vu, *Combinatorial problems in random matrix theory*, Proc. Int. Congr. Math. Vol. IV, 2014. pp. 489–508.
56. E. Wigner, *Characteristic vectors of bordered matrices with infinite dimensions*, Ann. Math. **62** (1955), 548–564.
57. L. Zdeborová and F. Krzakala, *Statistical physics of inference: Thresholds and algorithms*, Adv. Phys. **65** (2016), 453–552.

How to cite this article: A. Coja-Oghlan, A. A. Ergür, P. Gao, S. Hetterich, and M. Rolvien, *The rank of sparse random matrices*, Random Struct. Algorithms. **62** (2023), 68–130. https://doi.org/10.1002/rsa.21085

APPENDIX A: PROOF OF LEMMA 1.11

Since $\mathbb{E}[\lambda^r] < \infty$, the event $\mathscr{M} = \left\{\max_{i\in[s]} \lambda_i \le n/\ln^9 n\right\}$ has probability

$$\mathbb{P}[\mathscr{M}] = 1 - o_n(1/n). \tag{A1}$$

Moreover, fixing a small enough $\eta = \eta(\delta) > 0$ and a large enough $L = L(\eta) > 0$ and setting $Q_j = \sum_{i\in[s]} \mathbf{1}\{\lambda_i = j\}$, we obtain from the Chernoff bound that $\mathbb{P}\left[\forall j \le L : |Q_j - s\mathbb{P}[\lambda = j]| > \sqrt{n}\ln n\right] = o_n(1/n)$. Hence, by Bayes' rule,

$$\mathbb{P}\left[\exists j \le L : |Q_j - s\mathbb{P}[\lambda = j]| > \sqrt{n}\ln n | \mathscr{M}\right] = o_n(1/n). \tag{A2}$$

In addition, let $\mathscr{H} = \left\{h \in \mathbb{N} : (1+\eta)^{h-1}L \le n/\ln^9 n\right\}$ and for $h \in \mathscr{H}$ let

$$R_h = \sum_{j\ge 1} Q_j \mathbf{1}\{L(1+\eta)^{h-1} < j \le L(1+\eta)^h \wedge n/\ln^9 n\},$$

$$\overline{R}_h = s\sum_{j\ge 1} \mathbb{P}[\lambda = j]\mathbf{1}\{L(1+\eta)^{h-1} < j \le L(1+\eta)^h \wedge n/\ln^9 n\}.$$

Then the Chernoff bound and Bayes' rule yield

$$\mathbb{P}\left[\exists h \in \mathscr{H} : \left|R_h - \overline{R}_h\right| > \eta\overline{R}_h + \ln^2 n | \mathscr{M}\right] = o_n(1/n). \tag{A3}$$

Finally, given $\mathscr{M}$ and $|Q_j - s\mathbb{P}[\lambda = j]| \le \sqrt{n}\ln n$ for all $j \le L$ and $\left|R_h - \overline{R}_h\right| \le \eta\overline{R}_h + \ln^2 n$ for all $h \in \mathscr{H}$, we obtain

$$\begin{aligned}\frac{1}{s}\sum_{i=1}^{s}\lambda_i &\le \sum_{j=1} jQ_j/s + \sum_{h\in\mathscr{H}}(1+\eta)^h LR_h/s\\ &= o_n(1) + \mathbb{E}[\lambda\mathbf{1}\{\lambda \le L\}] + \sum_{h\in\mathscr{H}}(1+\eta)^{h+1}(\overline{R}_h + (\ln^2 n))/s \le \mathbb{E}[\lambda\mathbf{1}] + \delta/2 + o_n(1).\end{aligned}$$

Similarly, $\frac{1}{s}\sum_{i=1}^{s}\lambda_i \ge \mathbb{E}[\lambda\mathbf{1}] - \delta/2 + o_n(1)$. Thus, the assertion follows from (A1)–(A3)

APPENDIX B: STOCHASTIC VERSUS LINEAR INDEPENDENCE

A precursor of Proposition 2.4 for finite field was obtained in [7, lemma 3.1]. Instead of dealing with linear independence, that statement dealt with stochastic dependencies. Formally, given an $m\times n$-matrix A over a finite field $\mathbb{F}$, let μ_A be the probability distribution on $\mathbb{F}^n$ defined by

$$\mu_A(\sigma) = \mathbf{1}\{\sigma \in \ker A\}/|\ker A|.$$

(This definition is nonsensical over infinite fields for the obvious reason that $|\ker A| \in \{1, \infty\}$.) Let $\boldsymbol{\sigma} = \boldsymbol{\sigma}_A \in \mathbb{F}^n$ denote a sample from μ_A. The stochastic independence statement reads as follows.

Lemma B.1 ([7, lemma 3.1])**.** *For any $\delta > 0$, $\ell > 0$ and for any finite field $\mathbb{F}$ there exists $\mathscr{T} = \mathscr{T}(\delta, \ell, \mathbb{F}) > 0$ such that for any matrix A over $\mathbb{F}$ the following is true. Choose $\boldsymbol{\theta} \in [\mathscr{T}]$ uniformly at random. Then with probability at least $1 - \delta$ the matrix $A[\boldsymbol{\theta}]$ satisfies*

$$\sum_{I\subseteq[n]:|I|=\ell}\max_{\tau\in\mathbb{F}^I}\left|\mu_{A[\theta]}(\{\forall i \in I : \boldsymbol{\sigma}_i = \tau_i\}) - \prod_{i\in I}\mu_{A[\theta]}(\{\boldsymbol{\sigma}_i = \tau_i\})\right| < \delta n^{\ell}. \tag{B1}$$

In words, for most sets I of ℓ coordinates the joint distribution of the coordinates $(\boldsymbol{\sigma}_i)_{i\in I}$ is close to a product distribution in total variation distance. Furthermore, the number $\boldsymbol{\theta}$ of rows that we add to A is bounded in terms of ε, ℓ only; that is, $\boldsymbol{\theta}$ does not depend on the size $m \times n$ of A or on the matrix A itself. Lemma B.1 and its proof are inspired by the "pinning lemma" from [17].

The following lemma shows that how Proposition 2.4 implies Lemma B.1; in a nutshell, the lemma states that linear independence is stronger than stochastic independence.

Lemma B.2. *Let A be an $m \times n$-matrix over a finite field $\mathbb{F}$. Unless $I \subseteq [n]$ is a proper relation of A we have*

$$\mu_A(\{\forall i \in I : \boldsymbol{\sigma}_i = \tau_i\}) = \prod_{i\in I} \mu_A(\{\boldsymbol{\sigma}_i = \tau_i\}) \quad \text{for all } \tau \in \mathbb{F}^I. \tag{B2}$$

Proof. Since for every $\tau \in \mathbb{F}^I$ we have

$$\mu_A(\{\forall i \in I : \boldsymbol{\sigma}_i = \tau_i\}) = \mathbf{1}\{\forall i \in I \cap \mathfrak{F}(A) : \tau_i = 0\}\, \mu_A(\{\forall i \in I \setminus \mathfrak{F}(A) : \boldsymbol{\sigma}_i = \tau_i\}),$$
$$\prod_{i\in I} \mu_A(\{\boldsymbol{\sigma}_i = \tau_i\}) = \mathbf{1}\{\forall i \in I \cap \mathfrak{F}(A) : \tau_i = 0\} \prod_{i\in I\setminus\mathfrak{F}(A)} \mu_A(\{\boldsymbol{\sigma}_i = \tau_i\}),$$

we may assume that $I \cap \mathfrak{F}(A) = \emptyset$ by simply passing on to $I \setminus \mathfrak{F}(A)$ if necessary. Hence, the task reduces to proving (B2) under the assumption that $I \subseteq [n] \setminus \mathfrak{F}(A)$ is no relation of A.

To prove this statement let $N = \mathrm{nul}(A)$ and suppose that $\xi_1, \ldots, \xi_N \in \mathbb{F}^n$ form a basis of $\ker A$. Let $\Xi \in \mathbb{F}^{n\times N}$ be the matrix with columns $\xi_1, \ldots, \xi_N$ and let $\Xi_1, \ldots, \Xi_N$ signify the rows of Ξ. The homomorphism $z \in \mathbb{F}^N \to \ker A$, $z \mapsto \Xi z$ maps the uniform distribution on $\mathbb{F}^N$ to the uniform distribution μ_A on $\ker A$. Therefore, to prove (B2) it suffices to prove that the projection of this homomorphism to the I-rows, i.e., the map $z \in \mathbb{F}^N \mapsto (\Xi_i z)_{i\in I}$ is surjective. Equivalently, we need to show that

$$\mathrm{rk}\,(\Xi_i)_{i\in I} = |I|. \tag{B3}$$

Assume for contradiction that (B3) is violated. Then there exists a vector $z \in \mathbb{F}^I \setminus \{0\}$ such that $\sum_{i\in I} z_i \Xi_i = 0$. This implies that for all $x \in \mathbb{F}^n$,

$$Ax = 0 \quad \Rightarrow \quad \sum_{i\in I} z_i x_i = 0.$$

As a consequence, there exists a row vector y of length m such that $(yA)_j = \mathbf{1}\{i \in I\} z_i$ for all $j \in [n]$. Hence, $\emptyset \neq \mathrm{supp}(yA) \subseteq I$. Thus I is a relation of A, in contradiction to our assumption that it is not. ■

Thus, Lemma B.1 is an immediate consequence of Proposition 2.4 and Lemma B.2. Indeed, the proof of Proposition 2.4 renders the explicit bound $\mathscr{T} = \lceil 4\ell^3/\delta^4 \rceil + 1$ on the number of coordinates that need to get pegged. By comparison, the stochastic approach via the arguments from 7, 10 leads to a value of $\mathscr{T}$ that is exponential in ℓ (although it may be possible to improve this estimate via probabilistic arguments).

APPENDIX C: A SELF-CONTAINED PROOF OF THE UPPER BOUND ON THE RANK

The "$\leq$"-inequality in (1.3) was previously proved by Lelarge [41], who derived the bound from the Leibniz determinant formula and the formula for the matching number of random bipartite graphs

from [13]. The proof of that formula, however, is far from straightforward. Therefore, as a point of interest in this section we show that another idea from mathematical physics, the interpolation method from spin glass theory 25, 30, can be harnessed to obtain a self-contained proof of the upper bound on the rank. The proof uses similar ideas as the proof of the lower bound outlined in Section 2. Thus, phrased in terms of the nullity, the aim in this section is to show that a.a.s.

$$\mathrm{nul}(\boldsymbol{A})/n \geq \max_{\alpha\in[0,1]} \Phi(\alpha) + o_n(1). \tag{C1}$$

C.1 | The interpolation method

The basic idea behind the interpolation method is to construct a family of random matrices $\boldsymbol{A}_\varepsilon(t)$ parametrised by "time" t. At $t = \boldsymbol{m}_{\varepsilon,n}$ we obtain precisely the matrix $\boldsymbol{A}_{\varepsilon,n}$. At the other extreme, $\boldsymbol{A}_\varepsilon(0)$ is a block diagonal matrix whose nullity can be read off easily. To establish the lower bound we will control the change of the nullity with respect to t. By comparison to applications of the interpolation method to other combinatorial problems (e.g., 11, 17, 25, 49), the construction here is relatively elegant. In particular, throughout the interpolation we will be dealing with an actual random matrix, rather than some other, more contrived object.

Getting down to the details, apart from t and ε we need two further parameters: an integer $\mathscr{T} = \mathscr{T}(\varepsilon) \geq 0$ and a real $\beta \in [0,1]$, which, in order to obtain the optimal bound, we choose such that

$$\Phi(\beta) = \max_{\alpha\in[0,1]} \Phi(\alpha). \tag{C2}$$

Further, let $\boldsymbol{m}_{\varepsilon,n} \sim \mathrm{Po}((1-\varepsilon)dn/k)$. Also let $(\boldsymbol{k}_i, \boldsymbol{k}_i', \boldsymbol{k}_i'')_{i\geq 1}$ and $(\boldsymbol{d}_i)_{i\geq 1}$ be copies of $\boldsymbol{k}$ and $\boldsymbol{d}$, respectively, mutually independent and independent of $\boldsymbol{m}_{\varepsilon,n}$. Additionally, choose $\boldsymbol{\theta} \in [\mathscr{T}]$ uniformly and independently of everything else. Finally, recall that $(\boldsymbol{\zeta}_i, \boldsymbol{\xi}_i)_{i\geq 1}$ are uniformly distributed on the unit interval and independent of all other randomness.

The Tanner graph $\boldsymbol{G}_\varepsilon(t)$ has variable nodes

$$x_1, \ldots, x_n \quad \text{and} \quad (x_{i,j,h})_{i\in[\boldsymbol{m}_{\varepsilon,n}-t],\, j\in[\boldsymbol{k}_i'],\, h\in[\boldsymbol{k}_i'-1]}.$$

Moreover, let $\mathscr{F}_t$ be a random set that contains each of the variable nodes $x_{i,j,h}$ with probability β independently. Then the check nodes are

$$a_1, \ldots, a_t, \quad (b_{i,j})_{i\in[\boldsymbol{m}_{\varepsilon,n}-t],\, j\in[\boldsymbol{k}_i']}, \quad p_1, \ldots, p_{\boldsymbol{\theta}}, \quad f_{i,j,h} \quad \text{for each } x_{i,j,h} \in \mathscr{F}_t.$$

To define the edges of the Tanner graph let $\boldsymbol{\Gamma}_\varepsilon(t)$ be a random maximal matching of the complete bipartite graph with vertex sets

$$\bigcup_{i=1}^{n} \{x_i\} \times [\boldsymbol{d}_i], \quad \left(\bigcup_{i=1}^{t} \{a_i\} \times [\boldsymbol{k}_i]\right) \cup \left\{b_{i,j} : i \in [\boldsymbol{m}_{\varepsilon,n} - t],\, j \in [\boldsymbol{k}_i']\right\}.$$

For each matching edge $\{(x_i, s), (a_j, t)\} \in \boldsymbol{\Gamma}_\varepsilon(t)$ insert an edge between x_i and a_j into the Tanner graph and for each $\{(x_i, s), b_{j,h}\} \in \boldsymbol{\Gamma}_\varepsilon(t)$ insert an edge between x_i and $b_{j,h}$. Thus, $\boldsymbol{G}_\varepsilon(t)$ may contain multi-edges. Further, add an edge between x_i and p_i for $i = 1, \ldots, \boldsymbol{\theta}$ and add an edge between $x_{i,j,h}$ and $b_{i,j}$ for each $h \in [\boldsymbol{k}_i' - 1]$ as well as an edge between every $x_{i,j,h} \in \mathscr{F}_t$ and the check $f_{i,j,h}$. Finally, let $\boldsymbol{A}_\varepsilon(t)$ be the random matrix induced by $\boldsymbol{G}_\varepsilon(t)$. Formally, with the rows indexed by the check nodes and the columns indexed by the variable nodes, we let

$$(\boldsymbol{A}_\varepsilon(t))_{p_i,x_j} = \mathbf{1}\{i = j\} \qquad (i \in [\boldsymbol{\theta}], j \in [n]),$$

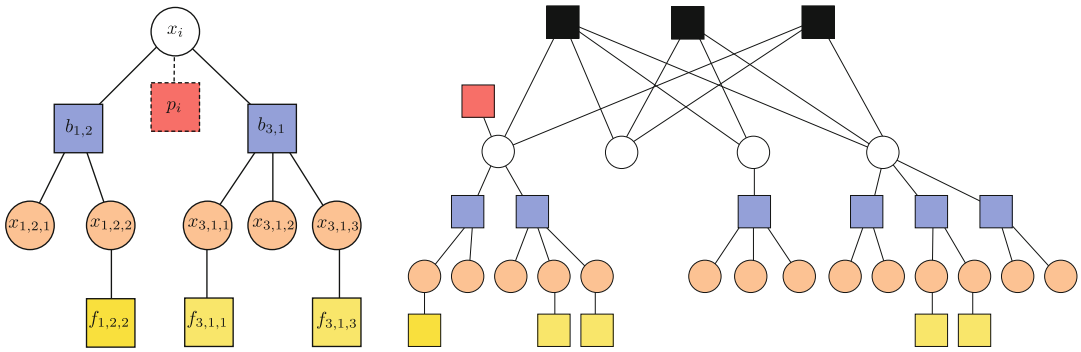

FIGURE C1 Left: sketch of the component of x_i at $t = 0$; the check p_i is present iff $i \le \boldsymbol{\theta}$. Right: sketch of the factor graph $\boldsymbol{G}_\varepsilon(t)$ for $0 < t < \boldsymbol{m}_{\varepsilon,n}$, with the $a_{i,j}$ coloured black and the other colours as in the left figure

$$
\begin{aligned}
(\boldsymbol{A}_\varepsilon(t))_{a_i,x_j} &= \chi_{\zeta_i,\xi_j} \sum_{u=1}^{\boldsymbol{k}_i} \sum_{v=1}^{\boldsymbol{d}_j} \mathbf{1}\left\{\{(x_j,v),(a_i,u) \in \boldsymbol{\Gamma}_\varepsilon(t)\}\right\} && (i \in [t], j \in [n]),\\
(\boldsymbol{A}_\varepsilon(t))_{b_{h,i},x_j} &= \mathbf{1}\left\{x_j \in \partial_{\boldsymbol{G}_\varepsilon(t)} b_{h,i}\right\} && (h \in [\boldsymbol{m}_{\varepsilon,n} - t], j \in [n]),\\
(\boldsymbol{A}_\varepsilon(t))_{b_{h,i},x_{u,v,w}} &= \mathbf{1}\{h = u,\ i = v\} && (h,u \in [\boldsymbol{m}_{\varepsilon,n} - t], i \in [\boldsymbol{k}'_h], v \in [\boldsymbol{k}'_u],\\
& && \quad w \in [\boldsymbol{k}'_u - 1]),\\
(\boldsymbol{A}_\varepsilon(t))_{f_{h,i,j},x_{u,v,w}} &= \mathbf{1}\{(h,i,j) = (u,v,w)\} && (h,u \in [\boldsymbol{m}_{\varepsilon,n} - t], i \in [\boldsymbol{k}'_h], j \in [v\boldsymbol{k}'_h - 1],\\
& && \quad v \in [\boldsymbol{k}'_u], w \in [\boldsymbol{k}'_u - 1]).
\end{aligned}
$$

All other entries of $\boldsymbol{A}_\varepsilon(t)$ are equal to zero.

The semantics is as follows. The checks a_i will play exactly the same role as before, that is, each is adjacent to $\boldsymbol{k}_i$ of the variable nodes $x_1, \ldots, x_n$ a.a.s. By contrast, each $b_{i,j}$ is adjacent to precisely one of the variables $x_1, \ldots, x_n$. In addition, $b_{i,j}$ is adjacent to the $\boldsymbol{k}'_i - 1$ variable nodes $x_{i,j,h}$, $h \in [\boldsymbol{k}'_i - 1]$. These variable nodes, in turn, are adjacent only to $b_{i,j}$ and to $f_{i,j,h}$ if $x_{i,j,h} \in \mathscr{F}$. The checks $f_{i,j,h}$ are unary, that is, $f_{i,j,h}$ simply forces $x_{i,j,h}$ to take the value zero. Finally, each of the checks p_i is adjacent to x_i only, i.e., $p_1, \ldots, p_{\boldsymbol{\theta}}$ just freeze $x_1, \ldots, x_{\boldsymbol{\theta}}$.

For $t = 1$ the Tanner graph contains $\boldsymbol{m}_{\varepsilon,n} \sim \mathrm{Po}((1-\varepsilon)dn/k)$ 'real' checks a_i and none of the checks $b_{i,j}$ or $f_{i,j,h}$. In effect, $\boldsymbol{A}_\varepsilon(1)$ is distributed precisely as $\boldsymbol{A}_\varepsilon$ from Section 2.2. By contrast, at $t = 0$ there are no checks a_i involving several of the variables $x_1, \ldots, x_n$. As a consequence, the Tanner graph decomposes into n connected components, one for each of the x_i. In fact, each component is a tree comprising x_i, some of the checks $b_{j,h}$ and their proprietary variables $x_{j,h,s}$ along with possibly a check $f_{j,h,s}$ that freezes $x_{j,h,s}$ to zero. For $i \in [\boldsymbol{\theta}]$ there is a check p_i freezing x_i to zero as well. Thus, $\boldsymbol{A}_\varepsilon(0)$ is a block diagonal matrix consisting of n blocks, one for each component. In effect, the rank of $\boldsymbol{A}_\varepsilon(0)$ will be easy to compute. Finally, for $0 < t < 1$ we have a blend of the two extremal cases. There will be some checks a_i and some $b_{i,j}$ with their retainer variables and checks; see Figure C1.

We are going to trace the nullity of $\boldsymbol{A}_\varepsilon(t)$ as t increases. But since the newly introduced variables $x_{i,j,h}$ inflate the nullity, we subtract the "obvious" correction term to retain the same scale throughout the process. In addition, we need a correction term to make up for the greater total number of check nodes in $\boldsymbol{A}_\varepsilon(0)$ by comparison to $\boldsymbol{A}_\varepsilon(\boldsymbol{m}_{\varepsilon,n})$. Thus, let

$$
\mathscr{N}_t = \operatorname{nul} \boldsymbol{A}_\varepsilon(t) + |\mathscr{F}_t| - \sum_{i=1}^{\boldsymbol{m}_{\varepsilon,n}-t} \boldsymbol{k}'_i(\boldsymbol{k}'_i - 1), \quad \mathscr{Y}_t = \sum_{i=1}^{\boldsymbol{m}_{\varepsilon,n}} (\boldsymbol{k}_i - 1)(\beta^{\boldsymbol{k}_i} - 1).
$$

The following two statements summarise the interpolation argument. First, we compute $\mathbb{E}[\mathscr{N}_0]$.

Proposition C.1. *For any fixed* $\theta \geq 0$ *we have* $n^{-1}\mathbb{E}[\mathscr{N}_0] = D(1-K'(\beta)/k)+dK'(\beta)/k-d+o_{\varepsilon,n}(1)$.

The next proposition provides monotonicity.

Proposition C.2. *For any* $\varepsilon > 0$ *there exists* $\mathscr{T} = \mathscr{T}(\varepsilon) > 0$ *such that with probability* $1-o_n(1/n)$ *uniformly for all* $0 \leq t < \boldsymbol{m}_{\varepsilon,n}$ *we have* $\mathbb{E}[\mathscr{N}_{t+1}+\mathscr{Y}_{t+1}|\boldsymbol{m}_{\varepsilon,n}] \geq \mathbb{E}[\mathscr{N}_t+\mathscr{Y}_t|\boldsymbol{m}_{\varepsilon,n}]+o_{\varepsilon,n}(1)$.

As an immediate consequence of Propositions C.1 and C.2 we obtain the desired lower bound on the nullity.

Corollary C.3. *We have* $\frac{1}{n}\mathbb{E}[\mathrm{nul}(\boldsymbol{A}_\varepsilon)] \geq \max_{\alpha\in[0,1]}\Phi(\alpha)+o_{\varepsilon,n}(1)$.

Proof. Proposition A.2 implies that

$$\begin{aligned}\mathbb{E}[\mathrm{nul}\,\boldsymbol{A}_{\varepsilon,n}] &= \mathbb{E}[\mathrm{nul}\,\boldsymbol{A}_\varepsilon(\boldsymbol{m}_{\varepsilon,n})] = \mathbb{E}[\mathscr{N}_{\boldsymbol{m}_{\varepsilon,n}}] = \mathbb{E}[\mathscr{N}_{\boldsymbol{m}_{\varepsilon,n}}+\mathscr{Y}_{\boldsymbol{m}_{\varepsilon,n}}]-\mathbb{E}[\mathscr{Y}_{\boldsymbol{m}_{\varepsilon,n}}]\\ &\geq \mathbb{E}[\mathscr{N}_0+\mathscr{Y}_0]-\mathbb{E}[\mathscr{Y}_{\boldsymbol{m}_{\varepsilon,n}}]+o_\varepsilon(n) = \mathbb{E}[\mathscr{N}_0]-\mathbb{E}[\mathscr{Y}_{\boldsymbol{m}_{\varepsilon,n}}]+o_{\varepsilon,n}(n). \qquad \text{(C3)}\end{aligned}$$

Further, by Proposition C.1,

$$\begin{aligned}\frac{1}{n}\mathbb{E}[\mathscr{N}_0] &= -d+dK'(\beta)/k+D(1-K'(\beta)/k)+o_{\varepsilon,n}(1),\\ \frac{1}{n}\mathbb{E}[\mathscr{Y}_{\boldsymbol{m}_{\varepsilon,n}}] &= \frac{d}{k}\left(\beta K'(\beta)-k+1-K(\beta)\right)+o_{\varepsilon,n}(1).\end{aligned}$$

Hence, (C2) yields

$$n^{-1}(\mathbb{E}[\mathscr{N}_0]-\mathbb{E}[\mathscr{Y}_{\boldsymbol{m}_{\varepsilon,n}}]) = \Phi(\beta)+o_\varepsilon(1) = \max_{\alpha\in[0,1]}\Phi(\alpha)+o_{\varepsilon,n}(1),$$

and the assertion follows from (C3). ∎

Combining Proposition 2.6, Proposition 2.8 and Corollary C.3 and the standard concentration for $\mathrm{nul}\,\boldsymbol{A}_\varepsilon$ from Lemma 4.7 completes the proof of (C1). We proceed to prove Propositions C.1 and C.2.

C.2 | Proof of Proposition C.1

Each component of $\boldsymbol{G}_\varepsilon(0)$ contains precisely one of the variable nodes $x_1,\ldots,x_n$. In effect, $\boldsymbol{A}_\varepsilon(0)$ has a block diagonal structure, and the overall nullity is nothing but the sum of the nullities of the blocks. It therefore suffices to calculate the nullity of the block $\boldsymbol{B}_s$ representing the connected component of x_s. Indeed, because $\sum_{s=1}^n|\partial^2 x_s| = \sum_{i\leq \boldsymbol{m}'_\varepsilon(0)}\boldsymbol{k}'_i(\boldsymbol{k}'_i-1)$ and $\sum_{s=1}^n|\partial^2 x_s\cap\mathscr{F}_0| = |\mathscr{F}_0|$ we have

$$\mathscr{N}_0 = \sum_{s=1}^n \boldsymbol{N}_s, \qquad \text{where} \quad \boldsymbol{N}_s = \mathrm{nul}(\boldsymbol{B}_s)-\left|\partial^2 x_s\right|+\left|\partial^2 x_s\cap\mathscr{F}_0\right|.$$

Consequently, since $\boldsymbol{\theta} = O_n(1)$ it suffices to prove that

$$\mathbb{E}[\boldsymbol{N}_s] = \begin{cases} dK'(\beta)/k+D(1-K'(\beta)/k)-d+o_\varepsilon(1) & \text{if } s>\boldsymbol{\theta},\\ O_n(1) & \text{otherwise.}\end{cases} \qquad \text{(C4)}$$

In fact, the second case in (C4) simply follows from $\boldsymbol{N}_s \leq \boldsymbol{d}_s$ and $\mathbb{E}[\boldsymbol{d}_s] = O_n(1)$ for all s.

Hence, suppose that $s > \boldsymbol{\theta}$. As $|\boldsymbol{N}_s| \leq \boldsymbol{d}_s$ and $\mathbb{E}[\boldsymbol{d}_s^r] = O_{\varepsilon,n}(1)$ for an $r > 2$ we find $\xi > 0$ such that

$$\mathbb{E}[|\boldsymbol{N}_s|\mathbf{1}\{\boldsymbol{d}_s > \varepsilon^{\xi-1/2}\}] = o_{\varepsilon,n}(1). \tag{C5}$$

Moreover, let $\Xi = \sum_{i=1}^{\boldsymbol{m}'_\varepsilon(0)} \boldsymbol{k}'_i \mathbf{1}\left\{\boldsymbol{k}'_i > \varepsilon^{-8}\right\}$, $\boldsymbol{M}'_j = \sum_{i=1}^{\boldsymbol{m}'_\varepsilon(0)} \mathbf{1}\{\boldsymbol{k}'_i = j\}$. Because $\mathbb{E}[\boldsymbol{k}^2] = O_{\varepsilon,n}(1)$ we have

$$\mathbb{E}[\Xi] \leq \frac{dn}{k}\mathbb{E}\left[\boldsymbol{k}\mathbf{1}\{\boldsymbol{k} \geq \varepsilon^{-8}\}\right] = nO_{\varepsilon,n}(\varepsilon^8), \tag{C6}$$

while $\boldsymbol{M}'_j \sim (1-\varepsilon)dn\mathbb{P}[\boldsymbol{k} = j]/k$ for all $j \leq \varepsilon^{-8}$ a.a.s. by Chebyshev's inequality. Hence, introducing the event

$$\mathscr{E}_s = \Big\{\boldsymbol{d}_s \leq \varepsilon^{\xi-1/2},\ \Xi \leq n\varepsilon^6,\ \forall j \leq \varepsilon^{-8} : \boldsymbol{M}'_j \sim (1-\varepsilon)dn\mathbb{P}[\boldsymbol{k} = j]/k,\ \sum_{i=1}^{n} \boldsymbol{d}_i \sim dn,\ \sum_{i\geq 3} i\boldsymbol{M}'_i \sim (1-\varepsilon)dn\Big\},$$

we obtain from (C5) and (C6) that

$$\mathbb{E}[\boldsymbol{N}_s] = \mathbb{E}[\boldsymbol{N}_s\mathbf{1}_{\mathscr{E}_s}] + o_{\varepsilon,n}(1). \tag{C7}$$

With $\boldsymbol{\gamma} \leq \boldsymbol{d}_s$ the actual degree of x_s in $\boldsymbol{G}_\varepsilon(s)$, let $\boldsymbol{\kappa}_1, \ldots, \boldsymbol{\kappa}_{\boldsymbol{\gamma}}$ be the degrees of the checks adjacent to x_s. We claim that given $\mathscr{E}_s$ and $\boldsymbol{d}_s$,

$$d_{\mathrm{TV}}((\boldsymbol{\kappa}_1, \ldots, \boldsymbol{\kappa}_{\boldsymbol{\gamma}}), (\hat{\boldsymbol{k}}_1, \ldots, \hat{\boldsymbol{k}}_{\boldsymbol{d}_s})) = o_{\varepsilon,n}(\varepsilon^{1/2}). \tag{C8}$$

Indeed, on $\mathscr{E}_s$ the probability that x_s is adjacent to a check of degree greater than ε^{-8} is $O_{\varepsilon,n}(\boldsymbol{d}_s\Xi/\sum_{j\geq 3} j\boldsymbol{M}'_j) = o_{\varepsilon,n}(\varepsilon)$. Further, given $\mathscr{E}_s$ we have

$$\sum_{j\geq 3} j\boldsymbol{M}'_j \geq (1-2\varepsilon)dn,$$

and thus $\mathbb{P}[\boldsymbol{\gamma} < \boldsymbol{d}_s|\mathscr{E}_s] = o_{\varepsilon,n}(\varepsilon^{1/2})$. Moreover, given $\boldsymbol{\gamma} = \boldsymbol{d}_s$, for each $i \in [\boldsymbol{d}_s]$ the probability that the ith clone of x_s gets matched to a check of degree $j \leq \varepsilon^{-8}$ is

$$j\boldsymbol{M}'_j/\sum_{h\geq 3} h\boldsymbol{M}'_h = j\mathbb{P}\left[\boldsymbol{k} = j\right]/k + o_n(1) = \mathbb{P}\left[\hat{\boldsymbol{k}} = j\right] + o_n(1).$$

These events are asymptotically independent for the different clones. Thus, we obtain (C8).

Finally, we can easily compute $\boldsymbol{N}_s$ given the vector $(\boldsymbol{\kappa}_1, \ldots, \boldsymbol{\kappa}_{\boldsymbol{\gamma}})$. The matrix $\boldsymbol{B}_s$ has fairly simple structure. The first $\boldsymbol{\gamma}$ rows have a non-zero entry in the first column representing x_s. Additionally, for $i = 1, \ldots, \boldsymbol{\gamma}$ the ith row contains $\boldsymbol{\kappa}_i - 1$ further non-zero entries, and the columns where theses non-zero entries occur are disjoint for all i. Finally, at the bottom of the matrix there is a block freezing the variables in $\mathscr{F}_0 \cap \partial^2 x_s$ to zero. We therefore claim that the rank of the matrix works out to be

$$\mathbb{E}[\mathrm{rk}(\boldsymbol{B}_s)|\boldsymbol{\kappa}_1, \ldots, \boldsymbol{\kappa}_{\boldsymbol{\gamma}}] = \sum_{i=1}^{\boldsymbol{\gamma}}(1-\beta^{\boldsymbol{\kappa}_i-1}) + |\mathscr{F}_0 \cap \partial^2 x_s| + 1 - \prod_{i=1}^{\boldsymbol{\gamma}}(1-\beta^{\boldsymbol{\kappa}_i-1}). \tag{C9}$$

To see this, let us first compute the rank of the matrix $\boldsymbol{B}_s'$ without the first column. Then row $i \in [\boldsymbol{\gamma}]$ contributes to the rank unless all the variables in the corresponding equation other than x_s belong to $\mathscr{F}_0$, an event that occurs with probability $\beta^{\boldsymbol{\kappa}_i - 1}$; hence the first summand. In addition, the $|\mathscr{F}_0 \cap \partial^2 x_s|$ rows pegging variables to zero contribute to the rank (second summand). Furthermore, going back to $\boldsymbol{B}_s$, the first column adds to the rank unless none of the first $\boldsymbol{\gamma}$ rows of $\boldsymbol{B}_s'$ gets zeroed out completely, an event that has probability $\prod_{i=1}^{\boldsymbol{\gamma}}(1-\beta^{\boldsymbol{\kappa}_i-1})$. Since

$$\begin{aligned}\mathbb{E}[\boldsymbol{N}_s|\boldsymbol{\kappa}_1, \dots, \boldsymbol{\kappa}_{\boldsymbol{\gamma}}] &= 1 + \sum_{i=1}^{\boldsymbol{\gamma}}(\boldsymbol{\kappa}_i - 1) - \mathbb{E}[\mathrm{rk}(\boldsymbol{B}_s)|\boldsymbol{\kappa}_1, \dots, \boldsymbol{\kappa}_{\boldsymbol{\gamma}}] - \mathbb{E}\left[\left|\partial^2 x_s\right| - \left|\partial^2 x_s \cap \mathscr{F}_0\right| |\boldsymbol{\kappa}_1, \dots, \boldsymbol{\kappa}_{\boldsymbol{\gamma}}\right] \\ &= 1 - \mathbb{E}[\mathrm{rk}(\boldsymbol{B}_s)|\boldsymbol{\kappa}_1, \dots, \boldsymbol{\kappa}_{\boldsymbol{\gamma}}] + \mathbb{E}\left[\left|\partial^2 x_s \cap \mathscr{F}_0\right| |\boldsymbol{\kappa}_1, \dots, \boldsymbol{\kappa}_{\boldsymbol{\gamma}}\right],\end{aligned}$$

substituting (C9) in yields

$$\mathbb{E}[\boldsymbol{N}_s|\boldsymbol{\kappa}_1, \dots, \boldsymbol{\kappa}_{\boldsymbol{\gamma}}] = \prod_{i=1}^{\boldsymbol{\gamma}}(1-\beta^{\boldsymbol{\kappa}_i-1}) - \sum_{i=1}^{\boldsymbol{\gamma}}(1-\beta^{\boldsymbol{\kappa}_i-1}). \tag{C10}$$

Combining (C7), (C8) and (C10) completes the proof.

C.3 | Proof of Proposition C.2

To couple the random variables $\mathscr{N}_{t+1}$ and $\mathscr{N}_t$ we need to investigate short linear relations among the cavities, that is, the clones from $\bigcup_{i=1}^n \{x_i\} \times [\boldsymbol{d}_i]$ that are not incident to an edge of $\boldsymbol{\Gamma}_\varepsilon(t)$. Denote this set by $\mathscr{C}(t)$. Further, let P_t be the distribution on the set of variables induced by drawing a random cavity, i.e.,

$$P_t(x_i) = |\mathscr{C}(t) \cap (\{x_i\} \times [\boldsymbol{d}_i])|/|\mathscr{C}(t)|,$$

and let $\boldsymbol{y}_1, \boldsymbol{y}_2 \dots$ be independent samples from P_t.

Lemma C.4. *For any $\delta > 0$ and $\ell > 0$ there is $\mathscr{T} = \mathscr{T}(\delta, \ell) > 0$ such that*

$$\mathbb{P}\left[\boldsymbol{y}_1, \dots, \boldsymbol{y}_\ell \text{ form a proper relation}\right] < \delta.$$

Proof. The choice of $\boldsymbol{m}_{\varepsilon,n}$ guarantees that $|\mathscr{C}(t)| \geq \varepsilon n/2$ a.a.s. Moreover, since $\mathbb{E}[\boldsymbol{d}] = O_{\varepsilon,n}(1)$ we find $L = L(\varepsilon, \delta) > 0$ such that the event $\mathscr{L} = \left\{\sum_{i=1}^n \boldsymbol{d}_i \mathbf{1}\{\boldsymbol{d}_i > L\} < \varepsilon\delta^2 n/16\right\}$ has probability $\mathbb{P}[\mathscr{L}] \geq 1 - \delta/8$. Therefore, we may condition on $\mathscr{E} = \mathscr{L} \cap \{|\mathscr{C}(t)| \geq \varepsilon n/2\}$.

Let $\boldsymbol{x}_1, \dots, \boldsymbol{x}_\ell$ be variables drawn uniformly with replacement from $V_n = \{x_1, \dots, x_n\}$. Then on the event $\mathscr{E}$ we have, for any ℓ-tuple $y_1, \dots, y_\ell$ of variables,

$$\mathbb{P}\left[\boldsymbol{y}_1 = y_1, \dots, \boldsymbol{y}_\ell = y_\ell | \boldsymbol{A}_\varepsilon(t)\right] \leq \mathbb{P}\left[\boldsymbol{x}_1 = y_1, \dots, \boldsymbol{x}_\ell = y_\ell | \boldsymbol{A}_\varepsilon(t)\right] (2L/\varepsilon)^\ell + \delta^2.$$

Consequently, because the distribution of $\boldsymbol{G}_\varepsilon(t) - \{p_1, \dots, p_\theta\}$ is invariant under permutations of $x_1, \dots, x_n$, Remark 3.6 shows that $\mathbb{P}\left[\boldsymbol{x}_1 = y_1, \dots, \boldsymbol{x}_\ell = y_\ell | \boldsymbol{A}_\varepsilon(t)\right] < \delta(\varepsilon/(2L))^\ell/2$, provided that $\mathscr{T} = \mathscr{T}(\delta, \ell)$ is large enough. ∎

We proceed to derive Proposition C.2 from Lemma C.4 and a coupling argument. Let $\boldsymbol{G}_\varepsilon'(t)$ be the Tanner graph obtained from $\boldsymbol{G}_\varepsilon(t+1)$ by removing the check a_{t+1}, let $\boldsymbol{A}_\varepsilon'(t)$ be the corresponding matrix and let

$$\mathscr{N}_t' = \mathrm{nul}\boldsymbol{A}_\varepsilon'(t) + |\mathscr{F}_{t+1}| - \sum_{i=1}^{\boldsymbol{m}_{\varepsilon,n}-t-1} \boldsymbol{k}_i'(\boldsymbol{k}_i' - 1).$$

Then clearly

$$\mathbb{E}\left[\mathscr{N}_{t+1} - \mathscr{N}_t | \boldsymbol{m}_{\varepsilon,n}\right] = \mathbb{E}\left[\mathscr{N}_{t+1} - \mathscr{N}_t' | \boldsymbol{m}_{\varepsilon,n}\right] - \mathbb{E}\left[\mathscr{N}_t - \mathscr{N}_t' | \boldsymbol{m}_{\varepsilon,n}\right].$$

Let $\boldsymbol{\alpha} \in [0,1]$ be the fraction of frozen cavities in $\boldsymbol{G}_\varepsilon'(t)$, with the convention that $\boldsymbol{\alpha} = 0$ if the set $\mathscr{C}'(t)$ of these cavities is empty.

Lemma C.5. *We have* $\mathbb{E}\left|\mathbb{E}[\mathscr{N}_{t+1} - \mathscr{N}_t' | \boldsymbol{A}_\varepsilon(t)', \boldsymbol{m}_{\varepsilon,n}] - (K(\boldsymbol{\alpha}) - 1)\right| = o_{\varepsilon,n}(1)$.

Proof. The random matrix $\boldsymbol{A}_\varepsilon(t+1)$ is obtained from $\boldsymbol{A}_\varepsilon'(t)$ by inserting a new random check a_{t+1}. Pick $\zeta = \zeta(\varepsilon) > 0$ small enough and $\delta = \delta(\zeta) > 0$ smaller still. Since $|\mathrm{nul}(\boldsymbol{A}_\varepsilon'(t)) - \mathrm{nul}(\boldsymbol{A}_\varepsilon(t+1))| \le 1$ and $\mathbb{E}[\boldsymbol{k}^2] = O_{\varepsilon,n}(1)$ we may condition on the event that $\boldsymbol{k}_{t+1} \le \varepsilon^{-1}$. Similarly, Lemma A.4 shows that we may assume that the set $\mathscr{X}$ of variables of $\boldsymbol{G}_\varepsilon'(t)$ where the new check node a_{t+1} attaches does not form a proper relation, provided that $\mathscr{T} = \mathscr{T}(\varepsilon)$ is chosen sufficiently large. Therefore, Lemma 2.5 yields

$$\begin{aligned}\mathbb{E}[\mathscr{N}_{t+1} - \mathscr{N}_t | \boldsymbol{A}_\varepsilon(t)', \boldsymbol{m}_{\varepsilon,n}] &= \mathbb{E}[\mathrm{nul}(\boldsymbol{A}_\varepsilon(t+1)) - \mathrm{nul}(\boldsymbol{A}_\varepsilon'(t)) | \boldsymbol{A}_\varepsilon'(t), \boldsymbol{m}_{\varepsilon,n}] \\ &= \mathbb{E}\left[\boldsymbol{\alpha}^{\boldsymbol{k}_{t+1}} - 1 | \boldsymbol{A}_\varepsilon'(t), \boldsymbol{m}_{\varepsilon,n}\right] + o_\varepsilon(1) = K(\boldsymbol{\alpha}) - 1 + o_{\varepsilon,n}(1),\end{aligned}$$

as claimed. ∎

Lemma C.6. *Let* $Q(\alpha,\beta) = \mathbb{E}\left[\boldsymbol{k}(\alpha\beta^{\boldsymbol{k}-1} - 1)\right]$ *for* $\alpha \in [0,1]$. *Then* $\mathbb{E}\left|\mathbb{E}\left[\mathscr{N}_t - \mathscr{N}_t' | \boldsymbol{A}_\varepsilon'(t), \boldsymbol{m}_{\varepsilon,n}\right] - Q(\boldsymbol{\alpha},\beta)\right| = o_{\varepsilon,n}(1)$.

Proof. The factor graph $\boldsymbol{G}_t(\varepsilon)$ is obtained from $\boldsymbol{G}_t'(\varepsilon)$ by adding the checks $b_{\boldsymbol{m}_{\varepsilon,n}-t-1,h}$ for $h \in [\boldsymbol{k}'_{\boldsymbol{m}_\varepsilon - t-1}]$, the corresponding variables $x_{\boldsymbol{m}_{\varepsilon,n}-t-1,h,j}$ and possibly their respective checks $f_{\boldsymbol{m}_{\varepsilon,n}-t-1,h,j}$. Since by construction

$$|\mathscr{N}_t - \mathscr{N}_t'| \le \boldsymbol{k}'_{\boldsymbol{m}_\varepsilon - t-1},$$

and $\mathbb{E}[\boldsymbol{k}^2] = O_{\varepsilon,n}(1)$ we may condition on the event that $\boldsymbol{k}'_{\boldsymbol{m}_\varepsilon - t-1} \le \varepsilon^{-1}$. In effect, Lemma C.4 shows that we may assume the set $\mathscr{X}$ of cavities adjacent to the new checks $b_{\boldsymbol{m}_{\varepsilon,n}-t-1,h}$ does not form a proper relation, provided that $\mathscr{T} = \mathscr{T}(\varepsilon)$ is chosen large enough. Moreover, the number of frozen cavities in $\mathscr{X}$ is within $o_n(1)$ of a binomial distribution $\mathrm{Bin}(\boldsymbol{k}'_{\boldsymbol{m}_{\varepsilon,n}-t-1}, \boldsymbol{\alpha})$ in total variation. Therefore, Lemma 2.5 shows that

$$\begin{aligned}\mathbb{E}\left[\mathscr{N}_t - \mathscr{N}_t' | \boldsymbol{A}_{\varepsilon,n}'(t), \boldsymbol{m}_{\varepsilon,n}\right] &= \mathbb{E}[\mathrm{nul}(\boldsymbol{A}_\varepsilon(t)) - \mathrm{nul}(\boldsymbol{A}_\varepsilon'(t)) - \boldsymbol{k}'_{\boldsymbol{m}_{\varepsilon,n}-t-1}(\boldsymbol{k}'_{\boldsymbol{m}_{\varepsilon,n}-t-1} - 1) + |\mathscr{F}'| | \boldsymbol{A}_\varepsilon'(t), \boldsymbol{m}_{\varepsilon,n}] \\ &= Q(\boldsymbol{\alpha},\beta) + o_{\varepsilon,n}(1),\end{aligned}$$

as claimed. ∎

Lemma C.7. *We have* $\mathbb{E}[\mathscr{Y}_{t+1} - \mathscr{Y}_t] = \mathbb{E}[(\boldsymbol{k}-1)(\beta^{\boldsymbol{k}} - 1)]$.

Proof. This is the result of a straightforward calculation. ∎

Proof of Proposition C.2. Combining Lemmas C.5–C.7, we obtain

$$\mathbb{E}[\mathscr{N}_{t+1} + \mathscr{Y}_{t+1}] - \mathbb{E}[\mathscr{N}_t + \mathscr{Y}_t] = \mathbb{E}\left[\boldsymbol{\alpha}^{\boldsymbol{k}} - 1 - \boldsymbol{k}(\boldsymbol{\alpha}\boldsymbol{\beta}^{\boldsymbol{k}-1} - 1) + (\boldsymbol{k}-1)(\boldsymbol{\beta}^{\boldsymbol{k}} - 1)\right] + o_{\varepsilon,n}(1). \quad \text{(C11)}$$

Since $x^k - kxy^{k-1} + (k-1)y^k \geq 0$ for all $k \geq 1$, $x, y \in [0,1]$, the assertion follows from (C11). ∎

APPENDIX D: Verification of $(\boldsymbol{m} - \boldsymbol{m}')/n$ and $\mathbf{rk}(\boldsymbol{A}')/n$

Let $\boldsymbol{m}_j$ denote the number of rows with exactly j nonzero entries. With standard concentration arguments, we know that a.a.s. $\boldsymbol{m}_0 \sim \boldsymbol{m}\mathbb{P}(\boldsymbol{k} = 0) \sim (dn/k)K(0)$, and $\boldsymbol{m}_1 \sim \boldsymbol{m}\mathbb{P}(\boldsymbol{k} = 1) \sim (dn/k)K'(0)$. Consequently, a.a.s. $(\boldsymbol{m} - \boldsymbol{m}')/n \sim d(1 - K(0) - K'(0))/k$.

For $\mathrm{rk}(\boldsymbol{A}')$, let X_i be the indicator variable that there exists a row with exactly one nonzero entry, and that nonzero entry occurs at the i-th column. Then $\mathrm{rk}(\boldsymbol{A}') = \sum_{i=1}^{n} X_i$. Conditioning on $\boldsymbol{m}_1$ and $\boldsymbol{D} = \sum_j j\boldsymbol{m}_j$, we know that $\mathbb{E}X_i = \sum_j \mathbb{P}(\boldsymbol{d} = j)(1 - (\boldsymbol{m}_1/\boldsymbol{D})^j)$ for every i. Since a.a.s. $\boldsymbol{m}_1/\boldsymbol{D} \sim K'(0)/k$, the standard concentration results immediately yield that a.a.s. $\mathrm{rk}(\boldsymbol{A}')/n \sim 1 - \sum_j \mathbb{P}(\boldsymbol{d} = j)(K'(0)/k)^j = 1 - D(1 - K'(0)/k)$.