THE RANK OF SPARSE RANDOM MATRICES

AMIN COJA-OGHLAN, ALPEREN A. ERGÜR, PU GAO, SAMUEL HETTERICH, MAURICE ROLVIEN

ABSTRACT. We determine the rank of a random matrix *A* over an arbitrary field with prescribed numbers of non-zero 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. *MSC*: 05C80, 60B20, 94B05

1. Introduction

1.1. **Background and motivation.** A renown physicist allegedly once quipped that mathematics, insofar as it is exact, does not bear on reality. Yet over the past decades man-made reality has incorporated increasingly sophisticated mathematical constructions. Modern error-correcting codes are a case in point. For instance, the codebook of a low-density parity check code ('ldpc code'), a type of code that has become a mainstay of modern communications standards, comprises the kernel of a sparse random matrix over a finite field drawn from a diligently chosen distribution [77]. Celebrated recent results establish that ldpc codes meet the Shannon bound on memoryless channels, i.e., that they are information-theoretically optimal. They even admit efficient decoding algorithms [41, 57]. Moreover, beyond the realm of coding theory, sparse random matrices play a prominent role in remarkably diverse areas of modern mathematics and its ramifications. To name but a few, apart from their classical role in probability and mathematical physics [66, 81, 82], prominent applications include the hashing problem in computer science [30], Ramanujan graphs [5, 16, 37], statistical inference [56] and the study of algorithms for constraint satisfaction problems [1, 44].

Despite the great interest in sparse random matrices certain fundamental questions remained open. Perhaps the most conspicuous one concerns the rank. Although this parameter was already studied in early contributions [7, 8, 54], there has been no comprehensive formula for the rank of random matrices whose number of nonzero entries is of the same order of magnitude as the number of rows. The present paper furnishes such a formula. To be precise, we will determine the rank of a sparse random matrix with prescribed row and column degrees (viz. number of non-zero entries). 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 [77]. Moreover, the rank is directly related to the rate of the ldpc code, defined as the nullity of the check matrix divided by the number of columns, arguably the code's most basic parameter.

Lelarge [58] noticed that an upper bound on the rank of a sparse random matrix follows from the result on the matching number of random bipartite graphs from [15]. He conjectured his bound to be tight for sparse random matrices over \mathbb{F}_2 . We prove this conjecture. In fact, we prove a 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 non-zero matrix entries. Thus, the rank is governed by the *location* of the non-zero entries.

The proof of the rank formula unearths an interesting connection to statistical physics. Indeed, Lelarge [58] observed that a sophisticated but mathematically non-rigorous physics approach called the 'cavity method' renders a wrong prediction as to the rank for certain degree distributions. This discrepancy merits attention because the cavity method has been brought to bear on a very wide range of real-world problems, ranging from the thermodynamics of glasses to machine learning [84]. 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 optimal solution to a variational problem. A priori, this variational problem asks to optimise a functional called the Bethe free entropy

1

Coja-Oghlan's research is supported by DFG CO 646/4. Ergür's research is supported by Einstein Foundation, Berlin. Gao's research is supported by ARC DE170100716 and ARC DP160100835. This submission combines the two preprints arXiv:1810.07390 and arXiv:1906.05757. An extended abstract of this work is due to appear in the proceedings of the 31st ACM-SIAM Symposium on Discrete Algorithms ('SODA 2020').

¹The derivation of this erroneous prediction was posed as an exercise in [67, Chapter 19].

over an infinite-dimensional space of probability measures. Such optimisation 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 will be able to dramatically simplify the variational problem, arriving at a humble one-dimensional optimisation 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. For certain degree distributions the result differs from those that surfaced in the experiments from [3] or the heuristic derivations from [67], hence the discrepancy between the physics predictions and mathematical reality. Apart from remedying this 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 2.3 below). We expect that this perturbation will find future applications.

We 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 $(\zeta_i, \xi_i)_{i \geq 1}$ be mutually independent uniformly distributed [0,1]-valued random variables. Moreover, let $d, k \geq 0$ be integer-valued random variables such that $0 < \mathbb{E}[d^r] + \mathbb{E}[k^r] < \infty$ for a real r > 2 and set $d = \mathbb{E}[d]$, $k = \mathbb{E}[k]$. Let n > 0 be an integer divisible by the greatest common divisor of the support of k and let $m \sim \text{Po}(dn/k)$ be independent of the ζ_i, ξ_i . Further, let $(d_i, k_i)_{i \geq 1}$ be copies of d, k, mutually independent and independent of m, ζ_i, ξ_i . Given

$$\sum_{i=1}^{n} d_i = \sum_{i=1}^{m} k_i, \tag{1.1}$$

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

$$A_{ij} = \mathbf{1}\{a_i x_j \in E(\mathbf{G})\} \cdot \chi_{\zeta_i, \xi_i}.$$

Thus, the *i*-th row of A contains precisely k_i non-zero entries and the *j*-th column contains precisely d_j non-zero entries. Moreover, the non-zero entries of A are drawn in the vein of an exchangeable array by evaluating the function χ at a random position (ζ_i, ξ_j) . Routine arguments show that A is well-defined for large enough n, i.e., (1.1) is satisfied and there exists a simple G with the desired degrees with positive probability; see Proposition 1.7 below. Adopting coding jargon, we call G the *Tanner graph* of A.

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

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

is continuous.

Theorem 1.1. For any d, k we have, uniformly for all χ ,

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

Theorem 1.1 establishes a substantially generalised version of Lelarge's rank conjecture [58]. The theorem covers a very general model of sparse random matrices. Indeed, since d, k have finite means the matrix A is sparse, i.e., the expected number of non-zero entries is O(n) as $n \to \infty$. Yet because the degree distributions are subject only to the modest condition $\mathbb{E}[d^r] + \mathbb{E}[k^r] < \infty$, the typical maximum number of non-zero entries per row or column may approach \sqrt{n} . Furthermore, the choice of the non-zero entries of the matrix by way of the measurable map χ , reminiscent of an exchangeable array, allows for rather general choices of non-zero matrix entries. Of course, an

immediate special case is the random matrix whose non-zero 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 $\operatorname{rk}(A)/n \le 1 - \max_{\alpha \in [0,1]} \Phi(\alpha) + o(1)$ as $n \to \infty$ w.h.p. was already derived in [58] from the Leibniz determinant formula and the formula for the matching number of a random bipartite graph from [15].³ Nonetheless, in the appendix we give an independent proof of the upper bound as well, which is more direct and shorter than the combination [15, 58].

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 G is the subgraph 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. Extending prior results that dealt with the degrees of all check nodes coinciding [24, 70], we compute the likely number of variable and check nodes in the 2-core. Since d, k have finite second moments and $\Phi'(0) \le 0$ while $\Phi'(1) \ge 0$, we can define

$$\rho = \max\{x \in [0,1] : \Phi'(x) = 0\}. \tag{1.4}$$

Further, let

$$\phi(\alpha) = 1 - \alpha - D' \left(1 - K'(\alpha)/k \right)/d \tag{1.5}$$

so that $\Phi'(\alpha) = dK''(\alpha)\phi(\alpha)/k$.

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

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

We observe that the expressions on the r.h.s. of (1.6) evaluate to zero if $\rho = 0$.

Theorem 1.2 yields an elementary upper bound on the rank of A. Indeed, upper bounding the rank is equivalent to lower bounding the nullity. Counting only solutions to Ax = 0 where $x_i = 0$ for all variables that belong to the 2-core G_* , we obtain $\operatorname{nul}(A) \ge n - n^* - (m - m^*)$. Invoking Theorem 1.2, we thus find, without further ado, that, as $n \to \infty$.

$$rk(A)/n \le 1 - \Phi(\rho) + o(1)$$
 w.h.p. (1.7)

Another elementary upper bound can be deduced as follows. Let A' be the matrix comprising the rows of A that contain at most one non-zero entry and let m' be the number of such rows. Then $\operatorname{rk}(A) \leq m - m' + \operatorname{rk}(A')$. Moreover, routine arguments reveal that $(m - m')/n \sim d(1 - K(0) - K'(0))/k$ and $\operatorname{rk}(A')/n \sim 1 - D(1 - K'(0)/k)$ w.h.p. Hence,

$$rk(A)/n \le 1 - \Phi(0) + o(1)$$
 w.h.p. (1.8)

Combining (1.7)–(1.8), we obtain the 2-core bound

$$rk(A)/n \le 1 - max\{\Phi(0), \Phi(\rho)\} + o(1)$$
 w.h.p. (1.9)

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

Theorem 1.3. Assume that

- (i) either $Var(\mathbf{d}) = 0$ or $\mathbf{d} \sim Po_{\geq \ell}(\lambda)$ for an integer $\ell \geq 0$ and $\lambda > 0$, and
- (ii) either $Var(\mathbf{k}) = 0$ or $\mathbf{k} \sim Po_{>\ell'}(\lambda')$ for an integer $\ell' \ge 0$ and $\lambda' > 0$.

Then

$$\lim_{n\to\infty} \operatorname{rk}(A)/n = 1 - \max\{\Phi(0), \Phi(\rho)\} \qquad in \ probability.$$

²To 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 sub-intervals 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 non-zero entries drawn from χ .

³While [58] only dealt with matrices over \mathbb{F}_2 , but the argument extends to other fields without further ado.

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

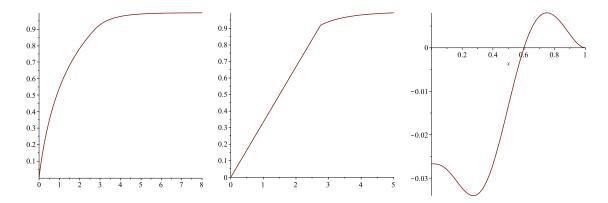


FIGURE 1. Left: the function $\Delta \mapsto 2 - \max_{\alpha \in [0,1]} \exp(-\Delta \exp(\Delta(\alpha-1))) + (1 + (1-\alpha)\Delta) \exp(\Delta(\alpha-1))$ for Example 1.4. Middle: the function $d \mapsto 1 - \max_{\alpha \in [0,1]} \exp(-d\alpha^{k-1}) - d(1-k\alpha^{k-1}+(k-1)\alpha^k)/k$ from Example 1.5 with k=3. Right: the function $\Phi(x)$ from Example 1.6.

On the basis of a canny but non-rigorous statistical physics approach called the cavity method several authors predicted that (over finite fields) the 2-core bound (1.9) is universally tight for all d, k. Alamino and Saad reached this conclusion by way of numerical experiments [3], while Mézard and Montanari [67] posed a non-rigorous but analytical derivation as an exercise. However, the prediction turns out to be erroneous. Indeed, Lelarge [58] produced an example of d, k whose function $\Phi(\alpha)$ attains its unique maximum at a value $0 < \alpha < \rho$. We will see another counterexample momentarily. On the positive side, Theorem 1.3 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 d, k and their resulting rank formulas.

Example 1.4 (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 $\operatorname{Po}(\Delta)$. Indeed, with the choice $d \sim \operatorname{Po}(\Delta)$ and $k \sim \operatorname{Po}(\Delta)$ the adjacency matrix $A(\mathbb{G}(n,n,p))$ and the random matrix A can be coupled such that $\operatorname{rk} A(\mathbb{G}(n,n,p)) = \operatorname{rk}(A) + o(n)$ w.h.p. Hence, Theorem 1.1 shows that over any field \mathbb{F} ,

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

in probability. Theorem 1.3 implies that the 2-core bound is tight in this example.

Example 1.5 (fixed row sums). Motivated by the minimum spanning tree problem on weighted random graphs, Cooper, Frieze and Pegden [26] studied the rank of the random matrix with degree distributions $k = k \ge 3$ fixed and $d \sim \text{Po}(d)$ over the field \mathbb{F}_2 . The same rank formula was obtained independently in [6] 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 non-zero entries is given by

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

Once more Theorem 1.3 shows that the 2-core bound is tight.

Example 1.6 (non-exact 2-core bound). There are plenty of choices of d, k where the 2-core bound fails to be tight, but degree distributions that render graphs 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 (1.4) being a local minimum of $\Phi(x)$. For instance, letting d, k be the distributions with $D(x) = (22x^2 + 3x^{11})/25$ and $K(x) = x^3$, we find $\rho = 1$ and $\Phi''(1) > 0$, while the global maximum is attained at $\alpha \approx 0.75$.

1.5. **Preliminaries.** Throughout the paper we consistently keep the assumptions on the distributions d, k listed in Section 1. In particular, $\mathbb{E}[d^r] + \mathbb{E}[k^r] < \infty$ for some r > 2. Because all-zero rows and columns do not add to the rank, replacing d, k by conditional random variables if necessary, we may assume that $d \ge 1$, $k \ge 1$. We write $\gcd(k)$ and $\gcd(d)$ for the greatest common divisor of the support of d and k, respectively. When working with d we tacitly assume that $\gcd(k)$ divides d. In order to highlight the number of columns we write d and d and d and d are d for the corresponding Tanner graph. The following proposition shows that d is well-defined.

Proposition 1.7. With probability $\Omega(n^{-1/2})$ over the choice of m, $(d_i)_{i\geq 1}$, $(k_i)_{i\geq 1}$ the condition (1.1) is satisfied and there exists a simple Tanner graph G with variable degrees d_1, \ldots, d_n and check degrees k_1, \ldots, k_m .

The proof of Proposition 1.7 is based on mildly intricate but ultimately routine arguments; we defer it to Section 4.2. We introduce the size-biased random variables

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

Throughout the paper we let $(k_i, d_i, \hat{k}_i, \hat{d}_i)_{i \ge 1}$ denote mutually independent copies of k, d, \hat{k}, \hat{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 multi-graphs. For instance, for a vertex v of a multi-graph G we denote by $\partial_G v$ the set of neighbours of v. More generally, for an integer $\ell \ge 1$ 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

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

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

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

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

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

Lemma 1.8. Let r > 2, $\delta > 0$ and suppose that $(\lambda_i)_{i \ge 1}$ are independent copies of a random variable $\lambda \ge 0$ 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 1.8 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 ' \geq '-part of (1.3), i.e., 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' [67]:

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 $X_n = \text{nul } A_n$. The coupling will be such that X_{n+1} is the rank of a random matrix obtained from 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 non-zero entries, an exceedingly complicated task. Two facts deliver us from this complexity. First, the positions of the non-zero entries of the new rows are (somewhat) random. Second, we develop a slight random permutation, applicable to any matrix, that diminishes the number of short linear relations (Proposition 2.3 below). In effect, the probability that there will be linear dependencies among the positions of the non-zero 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.** The following definition clarifies what we mean by 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 $\emptyset \neq supp(yA) \subseteq I$.
- 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 $\delta > 0$, $\ell \ge 1$ 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 non-trivial 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. Therefore, the i-th component of any vector in the kernel of A equals zero, in which case we call x_i 'frozen'. Further, excluding frozen variables, a proper relation I of A renders a non-trivial linear relation amongst 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 2.2. Let A be an $m \times n$ matrix and let $\theta \ge 0$ be an integer. Let $\mathbf{i}_1, \mathbf{i}_2, \dots, \mathbf{i}_{\theta} \in [n]$ 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 j-th new row has precisely one non-zero entry, namely a one in the \mathbf{i}_j -th column.

In other words, in $A[\theta]$ we expressly peg θ randomly chosen variables $x_{i_1}, \dots, x_{i_{\theta}}$ 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 2.3. For any $\delta > 0$, $\ell > 0$ there exists $\mathcal{T} = \mathcal{T}(\delta, \ell) > 0$ such that for any matrix A over any field \mathbb{F} the following is true. With $\theta \in [\mathcal{T}]$ chosen uniformly at random we have

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

The key feature of Proposition 2.3 is that the maximum number \mathcal{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 $|\operatorname{nul}(A) - \operatorname{nul} A[\boldsymbol{\theta}]| \leq \mathcal{T}$. Hence, while eliminating short proper relations, the perturbation does not shift the nullity significantly. Proposition 2.3 is a sweeping generalisation of a probabilistic result from [6], where the perturbation from Definition 2.2 was applied to matrices over finite fields to diminish stochastic dependencies amongst entries of randomly chosen vectors in the kernel. That argument, in turn, was inspired by ideas from information theory [22, 71, 76]. We will come back to this in Section 2.3.

We will incorporate the perturbation from Proposition 2.3 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 2.4. 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 non-zero columns of B. Moreover, obtain B_* from B by replacing for each $i \in I \cap \mathfrak{F}(A)$ the i-th column of B by zero. Unless I is a proper relation of A we have

$$\operatorname{nul}\begin{pmatrix} A & 0 \\ B & C \end{pmatrix} - \operatorname{nul} A = n' - \operatorname{rk}(B_* C). \tag{2.1}$$

To put Proposition 2.3 and Lemma 2.4 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 A_n from above. In line with the Aizenman-Sims-Starr scheme, a first stab at this problem might be to write a telescoping sum

$$\limsup_{n\to\infty} \frac{1}{n} \mathbb{E}[\operatorname{nul}(A_n)] = \limsup_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N-1} \mathbb{E}[\operatorname{nul}(A_{n+1})] - \mathbb{E}[\operatorname{nul}(A_n)] \le \limsup_{n\to\infty} \mathbb{E}[\operatorname{nul}(A_{n+1})] - \mathbb{E}[\operatorname{nul}(A_n)].$$

Then we should attempt to couple A_{n+1} and A_n so that we can write a single expectation

$$\mathbb{E}[\operatorname{nul}(A_{n+1})] - \mathbb{E}[\operatorname{nul}(A_n)] = \mathbb{E}[\operatorname{nul}(A_{n+1}) - \operatorname{nul}(A_n)]. \tag{2.2}$$

Ideally, to bring the tools from Section 2.1 to bear, under this coupling A_{n+1} should be obtained from A_n by adding one column and a few rows.

Unfortunately, this direct approach flounders for obvious reasons. For instance, depending on the distributions d, k, due to divisibility issues A_{n+1} may not even be defined for all n. 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 $\mathcal{T} = \mathcal{T}(\varepsilon) > 0$. Then for integers $n \geq \mathcal{T}$ we construct a random matrix $A_{\varepsilon,n}$ as follows. Like in Section 1.2 let $\chi : [0,1]^2 \to \mathbb{F}^*$ be a measurable map and let $(\zeta_i, \zeta_i)_{i \geq 1}$ be uniformly distributed [0,1]-valued random variables. Further, let

$$m_{\varepsilon,n} \sim \text{Po}((1-\varepsilon)dn/k)$$

Additionally, choose $\theta \in [\mathcal{T}]$ uniformly at random and, as before, let $(d_i)_{i\geq 1}$, $(k_i)_{i\geq 1}$ be copies of d, k. All of these random variables are mutually independent. Further, let $\Gamma_{\varepsilon,n}$ be a uniformly random maximal matching of the complete bipartite graph with vertex classes

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

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

$$(\boldsymbol{A}_{\varepsilon,n})_{p_i,x_j} = \mathbf{1}\left\{i = j\right\}$$

$$(i \in [\boldsymbol{\theta}], j \in [n]),$$

$$(\boldsymbol{A}_{\varepsilon,n})_{a_i,x_j} = \chi_{\boldsymbol{\zeta}_i,\boldsymbol{\xi}_j} \sum_{s=1}^{k_i} \sum_{t=1}^{d_j} \mathbf{1}\left\{\left\{(a_i,s),(x_j,t)\right\} \in \boldsymbol{\Gamma}_{\varepsilon,n}\right\}$$

$$(i \in [\boldsymbol{m}_{\varepsilon,n}], j \in [n]).$$

Morally, $A_{\varepsilon,n}$ mimics the matrix obtained from the original model 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_{θ} is to ensure that $A_{\varepsilon,n}$ is (δ, ℓ) -free for a small enough $\delta = \delta(\varepsilon)$ and a large enough $\ell = \ell(\varepsilon)$. Indeed, while Proposition 2.3 requires that a random set of θ variables be pegged, the checks p_1, \ldots, p_{θ} just freeze the first θ variables. But since the distribution of the Tanner graph $G_{\varepsilon,n} - \{p_1, \ldots, p_{\theta}\}$ is invariant under permutations of the variable nodes, both constructions are equivalent. The following proposition shows that going to $A_{\varepsilon,n}$ does not shift the rank significantly.

Proposition 2.5. For any any 0 < C < C' and any function $\mathcal{T} = \mathcal{T}(\varepsilon) \ge 0$ the following is true. If

$$\limsup_{\varepsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \mathbb{E}[\operatorname{nul}(A_{\varepsilon,n})] \le C \qquad then \qquad \lim_{n \to \infty} \mathbb{P}\left[\operatorname{nul}(A_n) \le C'n\right] = 1. \tag{2.3}$$

⁵For instance, suppose that d = 3 and k = 4 deterministically. Then (1.1) boils down to 4m = 3n, and thus A_n is well-defined only if n is divisible by four.

Analogously, if

$$\liminf_{\varepsilon \to 0} \liminf_{n \to \infty} \frac{1}{n} \mathbb{E}[\operatorname{nul}(A_{\varepsilon,n})] \geq C' \qquad then \qquad \lim_{n \to \infty} \mathbb{P}\left[\operatorname{nul}(A_n) \geq Cn\right] = 1.$$

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

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

most check nodes a_i have degree precisely k_i w.h.p. But we expect that about εdn variable nodes x_i will have degree less than d_i . In fact, w.h.p. $\Gamma_{\varepsilon,n}$ fails to cover about εdn 'clones' from the set $\bigcup_{i=1}^n \{x_i\} \times [d_i]$. Let us call such unmatched clones *cavities*.

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

$$\operatorname{nul} \mathbf{A}_{\varepsilon, n+1} = \operatorname{nul} \begin{pmatrix} \mathbf{A}_{\varepsilon, n} & 0 \\ \mathbf{B} & \mathbf{C} \end{pmatrix}, \tag{2.4}$$

where ${\it B}$ has n columns and ${\it d}_{n+1}$ rows and ${\it C}$ is a column vector of size ${\it d}_{n+1}$ w.h.p. But this direct attempt has a subtle flaw. Indeed, going from ${\it A}_{\varepsilon,n}$ to ${\it A}_{\varepsilon,n+1}$, (2.4) adds $\mathbb{E}[{\it d}_{n+1}]=d$ rows on the average. Yet actually we should be adding merely $\mathbb{E}[{\it m}_{\varepsilon,n+1}-{\it m}_{\varepsilon,n}]=(1-\varepsilon)d/k$ rows. To remedy this problem we borrow a trick from prior applications of the Aizenman-Sims-Starr scheme in combinatorics [6, 22, 23]. Namely, we set up a coupling under which both ${\it A}_{\varepsilon,n}, {\it A}_{\varepsilon,n+1}$ are obtained by adding a few rows/columns to a common 'base matrix' ${\it A}'$. Thus, instead of (2.4) we obtain

$$\operatorname{nul} \mathbf{A}_{\varepsilon,n} = \operatorname{nul} \begin{pmatrix} \mathbf{A}' & 0 \\ \mathbf{B}' & \mathbf{C}' \end{pmatrix}. \tag{2.5}$$

To be precise, C' above is a column vector with an expected $(1-\varepsilon)d$ non-zero entries and B, B' are matrices whose numbers of non-zero entries are bounded in expectation. Furthermore, the base matrix A' itself is quite similar to $A_{\varepsilon,n}$, except that A' has a slightly smaller number of rows. In Section 5 we will present the construction in full detail and apply Proposition 2.3 and Lemma 2.4 to prove the following upper bound on the change in nullity. Recall the function Φ from (1.2) and recall that $\Theta \in [\mathcal{F}]$ with $\mathcal{F} = \mathcal{F}(\varepsilon)$ dependent on ε only is the number of pinned variables in the construction of $A_{\varepsilon,n}$.

Proposition 2.6. There exists a function $\mathcal{T} = \mathcal{T}(\varepsilon) > 0$ such that

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

As an immediate consequence of Proposition 2.6 we obtain the desired upper bound on the nullity.

Corollary 2.7. We have

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

Proof. Proposition 2.6 yields

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

as claimed. \Box

Proof of Theorem 1.1. The desired lower bound on the rank of A_n is an immediate consequence of Proposition 2.5 and Corollary 2.7.

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 non-zero entries per row/column is bounded or unbounded. Subsequently we discuss the physicists' (non-rigorous) 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 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 [53] proved that such matrices are regular w.h.p. Vu [81] 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 [48, 79, 80], 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 [54, 55, 60, 61] 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 [38, 64] and the references therein.

A further line of work deals with random $m \times n$ matrices in which the number of non-zero entries per row diverges in the limit of large n but is of order $o_n(n)$. Relating the permanent and the determinant, Balakin [8] and, using delicate moment calculations, Blömer, Karp and Welzl [11] dealt with the rank of such matrices over finite fields. Moreover, using expansion arguments, Costello and Vu [27, 28] studied the real rank of random symmetric matrices of a similar density. They find that such matrices essentially have full rank w.h.p., apart from a small defect based on local phenomena. In the words of [28], "dependency [comes] from small configurations".

2.3.2. *Sparse matrices*. Matters are quite different in the sparse case where the average number of non-zero 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", i.e., by minimally linearly dependent sets of unbounded size.

The first major contribution dedicated to sparse matrices was a paper by Dubois and Mandler [31] 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 has full rank (i.e., equal to $m \wedge n$) w.h.p. Thus, the random matrix model is just the one from Example 1.5 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 $|\ker A|$, which boils down to a one-dimensional calculus problem. Matters get more complicated when one considers a greater number k>3 of non-zero entries per row. This more general problem, known as random k-XORSAT, was solved independently by Dietzfelbinger et al. [30] and by Pittel and Sorkin [75] 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 [42] managed to extend the method to \mathbb{F}_3 . However, extending this strategy to infinite fields is a non-starter as $|\ker A|$ may be infinite.

In a previous paper Ayre, Coja-Oghlan, Gao and Müller [6] applied the Aizenman-Sims-Starr scheme to the study of sparse random matrices with precisely k non-zero entries per row as in Example 1.5, 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 d, k rather than just the Poisson-constant degrees from Example 1.5, including degree sequences for which the 2-core bound fails to be tight (in contrast to Example 1.5). 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 generalisations of the models from [22].

Second, the rank calculation in [6] 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 σ of the kernel of the random matrix A. Specifically, [6, Lemma 3.1] analyses the impact of the perturbation from Definition 2.2 on a matrix $A \in \mathbb{F}^{m \times n}$ for a finite field \mathbb{F} . With $\sigma = (\sigma_1, \dots, \sigma_n) \in \ker(A)$ a uniformly random element of the kernel, the lemma shows that for a

large enough $\mathcal{T} = \mathcal{T}(\delta, \mathbb{F}) > 0$ and a uniformly random $0 \le \theta \le \mathcal{T}$,

$$\sum_{\substack{1 \le i < j \le n \\ \omega \ \omega' \in \mathbb{F}}} \mathbb{E} \left| \mathbb{P} \left[\boldsymbol{\sigma}_i = \omega, \boldsymbol{\sigma}_j = \omega' \mid A[\boldsymbol{\theta}] \right] - \mathbb{P} \left[\boldsymbol{\sigma}_i = \omega \mid A[\boldsymbol{\theta}] \right] \cdot \mathbb{P} \left[\boldsymbol{\sigma}_j = \omega' \mid A[\boldsymbol{\theta}] \right] \right| < \delta n^2, \tag{2.6}$$

As in Proposition 2.3, the necessary value of \mathcal{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 [9], (2.6) extends from pairwise independence to ℓ -wise independence, albeit with a weaker error bound δ . The result [6, Lemma 3.1] was inspired by general statements about probability measures on discrete cubes from [22, 71, 76].

Inherently, this stochastic approach does not generalise 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 summarised in Proposition 2.3, 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 2.3 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 (2.6) follows from the linear independence provided by Proposition 2.3, with a significantly improved bound on $\mathcal{T}(\delta)$. 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 [14], 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' [4]. 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 non-rigorous technique inspired by the statistical mechanics of disordered systems, it had been predicted erroneously that over finite fields the 2-core bound (1.9) on the rank of A is universally tight for general degree distributions d, k [3, 67]. 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 A over a finite field \mathbb{F}_q converges in probability to the solution of an optimisation 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 $(\mu_{i,j}^{\pi})_{i,j\geq 1}$ be a sequence of independent samples from π . Recalling (1.10), the *Bethe free entropy* is defined by

$$\mathcal{B}(\pi) = \mathbb{E}\left[\log_{q} \sum_{\sigma_{1} \in \mathbb{F}_{q}} \prod_{i=1}^{d} \sum_{\sigma_{2}, \dots, \sigma_{\hat{k}_{i}} \in \mathbb{F}_{q}} \mathbf{1} \left\{ \sum_{j=1}^{\hat{k}_{i}} \sigma_{j} \chi_{i, j} = 0 \right\} \prod_{j=2}^{\hat{k}_{i}} \boldsymbol{\mu}_{i, j}^{\pi}(\sigma_{j}) \right] - \frac{d}{k} \mathbb{E}\left[(\boldsymbol{k} - 1) \log_{q} \sum_{\sigma_{1}, \dots, \sigma_{k} \in \mathbb{F}_{q}} \mathbf{1} \left\{ \sum_{i=1}^{k} \sigma_{i} \chi_{1, i} = 0 \right\} \prod_{i=1}^{k} \boldsymbol{\mu}_{1, i}^{\pi}(\sigma_{i}) \right].$$
 (cf. [67, Chapter 14]).

The replica symmetric ansatz predicts that

$$\lim_{n \to \infty} \frac{1}{n} \operatorname{nul} A = \sup_{\pi \in \mathscr{P}^2(\mathbb{F}_q)} \mathscr{B}(\pi) \qquad \text{in probability.}$$
 (2.7)

For a detailed (heuristic) derivation of the Bethe free entropy and the prediction (2.7) we refer to [3]. But let us briefly comment on the intended semantics of π . Consider the Tanner graph G representing 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* $\mu_{A,x_j \to a_i}(\sigma)$ from x_j to a_i as follows. Obtain $A_{x_j \to a_i}$ from A by changing the ij-th matrix entry to zero; this corresponds to deleting the x_j - a_i -edge from the Tanner graph. Then $\mu_{A,x_j \to a_i}(\sigma)$ is the probability that in a uniformly random vector $\sigma \in \ker A_{x_j \to a_i}$ we have $\sigma_j = \sigma$. Further, define π_A as the empirical distribution of the $\mu_{A,x_j \to a_i}$ over the edges of the Tanner graph:

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

Then the replica symmetric ansatz predicts that π_A is asymptotically a maximiser of the Bethe free energy, i.e., $\sup_{\pi \in \mathscr{P}^2(\mathbb{F}_q)} \mathscr{B}(\pi) = \mathscr{B}(\pi_A) + o_n(1)$ w.h.p. Thus, the maximiser π in (2.7) is deemed to encode the Belief Propagation messages on the edges of the Tanner graph of 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 [6, Lemma 2.3]. Namely, any message $\mu_{A,x_j\to a_i}$ 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

$$\pi_{\alpha} = \alpha \delta_{\delta_0} + (1 - \alpha) \delta_{q^{-1} \mathbf{1}} \qquad (\alpha \in [0, 1]). \tag{2.8}$$

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

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

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

Alamino and Saad [3] tackled the optimisation problem (2.7) 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 (1.4); in fact, all their examples fall within the purview of Theorem 1.3.⁶ 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 1.6 and [58] provide counterexamples. The more sophisticated 1RSB cavity method is presented in [67, Chapter 19], where an exercise asks the reader to verify that the 2-core bound is tight (over finite fields). While Theorem 1.3 gives sufficient conditions for this to be correct, the aforementioned counterexamples apply.

2.4. **Organisation.** We proceed to prove Proposition 2.3, 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 2.5. Additionally, Section 4 contains Proposition 1.7, 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 2.6. Further, Section 6 deals with the proof of Theorem 1.2 and in Section 7 we prove Theorem 1.3. For the sake of completeness a proof of Lemma 1.8 is included in Appendix A. Moreover, in Appendix B we elaborate on the relation between the algebraic perturbation from Proposition 2.3 and the stochastic version from [6]. 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 2.3

In this section we prove Proposition 2.3 and Lemma 2.4. 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 2.3.** Choosing $\mathcal{T} = \mathcal{T}(\delta, \ell)$ sufficiently large, we may safely assume that the number $n > n_0(\delta, \ell)$ of columns of the matrix A is large enough. Moreover, 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 $\mathcal{R}_h(M)$ be the set of all minimal h-relations of M and set $R_h(M) = |\mathcal{R}_h(M)|$. Thus, $R_1(M) = |\mathfrak{F}(M)|$ is just the number of frozen variables of M. Additionally, let $\mathcal{R}_{\leq h}(M) = \bigcup_{1 \leq i \leq h} \mathcal{R}_i(M)$ and $R_{\leq h}(M) = |\mathcal{R}_{\leq h}(M)|$. Let $i_1, i_2, i_3, \ldots \in [n]$ be uniformly distributed independent random variables.

The proof of Proposition 2.3 is based on a potential function argument. To get started we observe that

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

This inequality implies that the random variable

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

⁶Strictly 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 $\Phi(\rho) > \Phi(0)$.

is non-negative. 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 [6] 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 $\mathbb{E}[\Delta_{\theta}] \leq \ell/\mathcal{T}$.

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

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

Summing this bound on r, we obtain

$$\sum_{\theta \in [\mathcal{T}]} \mathbb{E}[\Delta_{\theta}] \le \sum_{r=0}^{\ell-1} \sum_{j \ge 0} \mathbb{E}[\Delta_{r+j\ell}] \le \ell. \tag{3.2}$$

Since $\theta \in [\mathcal{T}]$ is chosen uniformly and independently of everything else, dividing (3.2) by \mathcal{T} completes the proof.

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 3.2.** 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

$$R_h(A[t]) < \delta n^h / \ell$$
 for all $2 \le h \le \ell$. (3.3)

Since every proper relation I of size $|I| = \ell$ contains a minimal h-relation $J \subseteq I$ for some $2 \le h \le \ell$, (3.3) implies that A[t] possesses fewer than δn^{ℓ} proper relations of size ℓ in total. Hence, if (3.3) 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 3.3. 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 $\mathcal{R}_{v,h}(A[t])$ be the set of all relations $I \in \mathcal{R}_h(A[t])$ that contain $v \in [n]$ and set $r_{v,t,h} = |\mathcal{R}_{v,h}(A[t])|$. Moreover, let $\mathcal{V}_{t,h}$ 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,

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

whence

$$|\mathcal{V}_{t,h}| \ge \frac{\delta hn}{2\ell}.\tag{3.4}$$

Consider $v \in \mathcal{V}_{t,h}$ along with a minimal h-relation $I \in \mathcal{R}_{v,h}(A[t])$. If $I = \{v, \boldsymbol{i}_{t+1}, \dots, \boldsymbol{i}_{t+h-1}\}$, i.e., 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 $\sup(yA[t]) = I$. Hence, if $I \setminus \{v\} = \{\boldsymbol{i}_{t+1}, \dots, \boldsymbol{i}_{t+h-1}\}$, then we can extend y to a row vector y' such that $\sup(y'A[t+\ell]) = \{v\}$, and thus $v \in \mathfrak{F}(A[t+h-1])$. Furthermore, since $(\boldsymbol{i}_{t+1}, \dots, \boldsymbol{i}_{t+h-1}) \in [n]^{h-1}$ is uniformly random, we conclude that

$$\mathbb{P}\left[I = \{v, \mathbf{i}_{t+1}, \dots, \mathbf{i}_{t+h-1}\} \mid A[t]\right] = (h-1)!/n^{h-1} \ge n^{1-h}.$$
(3.5)

Now, because every $v \in V_{t,h}$ satisfies $r_{v,t,h} \ge \delta h n^{h-1}/(2\ell)$, (3.5) implies that

$$\mathbb{P}\left[v \in \mathfrak{F}(A[t+h-1]) \mid A[t]\right] \ge r_{v,t,h}/n^{h-1} \ge \delta h/(2\ell). \tag{3.6}$$

We also notice that $V_{t,h} \cap \mathfrak{F}(A[t]) = \emptyset$ because no minimal h-relation contains a frozen variable. Therefore, combining (3.1), (3.4) and (3.6) and using linearity of expectation, we obtain

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

as desired.

Combining Lemmas 3.2 and 3.3, we immediately obtain the following.

Corollary 3.4. *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 2.3.

Proof of Proposition 2.3. We define $T = \{t \in [\mathcal{F}] : \mathbb{P}[A[t] \text{ fails to be } (\delta, \ell)\text{-free}] \geq \delta/2\}$ so that

$$\mathbb{P}\left[A[\boldsymbol{\theta}] \text{ is } (\delta, \ell)\text{-free}\right] > 1 - \delta/2 - \mathbb{P}\left[\boldsymbol{\theta} \in T\right]. \tag{3.7}$$

Hence, we are left to estimate $\mathbb{P}[\theta \in T]$. Applying Corollary 3.4, we obtain for every $t \in T$,

$$\mathbb{E}[\Delta_t] \ge \frac{\delta^2}{\ell^2} \cdot \mathbb{P}[A[t] \text{ fails to be } (\delta, \ell)\text{-free}] \ge \frac{\delta^3}{2\ell^2}.$$
 (3.8)

Moreover, averaging (3.8) on $t \in [\mathcal{T}]$ and applying Lemma 3.1, we obtain

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

Consequently, choosing $\mathcal{T} > 4\ell^3/\delta^4$ ensures $\mathbb{P}[\boldsymbol{\theta} \in T] \leq \delta/2$. Thus, the assertion follows from (3.7).

Remark 3.5. The proof presented in this section actually renders a slightly stronger statement than Proposition 2.3. Specifically, let A be an $m \times N$ -matrix and let $n \le N$. 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 $\theta = \theta(\delta, \ell)$ chosen as in Proposition 2.3 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 θ 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 2.4.** We are going to derive Lemma 2.4 from the following simpler statement.

Lemma 3.6. 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 non-zero 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

$$\operatorname{nul} A - \operatorname{nul} \begin{pmatrix} A \\ B \end{pmatrix} = \operatorname{rk}(B). \tag{3.9}$$

Writing $B_1, ..., B_{m'}$ for the rows of B and $r = \operatorname{rk}(B)$ for the rank and applying a row permutation if necessary, we may assume that $B_1, ..., B_r$ are linearly independent. Hence, to establish (3.9) it suffices to prove that for all $0 \le \ell < r$,

$$\operatorname{rk} \begin{pmatrix} A \\ B_1 \\ \vdots \\ B_{\ell} \end{pmatrix} < \operatorname{rk} \begin{pmatrix} A \\ B_1 \\ \vdots \\ B_{\ell+1} \end{pmatrix}. \tag{3.10}$$

In other words, we need to show that $B_{\ell+1}$ does not belong to the space spanned by B_1, \ldots, B_ℓ and the rows A_1, \ldots, 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 $0 \neq B_{\ell+1} - \sum_{i=1}^\ell x_i B_i = \sum_{i=m}^\ell y_i A_i$ and thus $\emptyset \neq \operatorname{supp} \sum_{i=m}^\ell y_i A_i \subseteq I$, in contradiction to the assumption that I is no relation of A. Hence, we obtain (3.10) and thus (3.9). Finally, to complete the proof of (2.1) we apply (3.9) to the matrices (A0) and (B0, obtaining

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

as desired.

Proof of Lemma 2.4. For any $i \in \mathfrak{F}(A)$ the i-th standard unit row vector can be written as a linear combination of the rows of A. Since elementary row operations do not alter the nullity of a matrix, we therefore find

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

The assertion thus follows from Lemma 3.6.

4. CONCENTRATION

The principal aim of this section is to prove Proposition 2.5, i.e., to argue that the rank of the actual matrix A_n that does not have any cavities and whose Tanner graph is simple is close to the expected rank of $A_{\varepsilon,n}$ w.h.p. In other words, we need to show that the rank of a random matrix is sufficiently concentrated that conditioning on

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

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 ([32, p. 130]). Suppose that $(X_i)_{i\geq 1}$ is a sequence of i.i.d. variables that take values in $\mathbb Z$ such that the greatest common divisor of the support of X_1 is one. Also assume that $\operatorname{Var}[X_1] = \sigma^2 \in (0, \infty)$. Then

$$\lim_{n\to\infty} \sup_{z\in\mathbb{Z}} \left| \sqrt{n} \mathbb{P} \left[\sum_{i=1}^n X_i = z \right] - \frac{\exp(-(z - n\mathbb{E}[X_1])^2/(2n\sigma^2))}{\sqrt{2\pi}\sigma} \right| = 0.$$

Subsequently, in Section 4.2 we calculate the probability of the event \mathcal{S} , proving Proposition 1.7 along the way. Finally, in Section 4.3 we complete the proof of Proposition 2.5.

4.1. **The event** \mathcal{D} . Because $\mathbb{E}[d^r] + \mathbb{E}[k^r] < \infty$ for an r > 2, the event

$$\mathcal{M} = \left\{ \max_{i \in [n]} d_i + \max_{i \in [m]} k_i \le \sqrt{n} / \log^9 n \right\} \quad \text{satisfies} \quad \mathbb{P}[\mathcal{M}] = 1 - o_n(1). \tag{4.1}$$

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

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

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

Second, suppose that Var(d) > 0 but Var(k) = 0. Then Theorem 4.1 and (4.1) show that

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

Further, given $|dn - \sum_{i=1}^{n} d_i| \le \sqrt{n}$ and given k divides $\sum_{i=1}^{n} d_i$, the event $km = \sum_{i=1}^{n} d_i$ has probability $\Theta_n(n^{-1/2})$ by the local limit theorem for the Poisson distribution.

The case that $Var(\boldsymbol{d}) = 0$ but $Var(\boldsymbol{k}) > 0$ 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

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

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

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

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

A similar argument applies in the final case Var(d) > 0, Var(k) > 0. Indeed, Theorem 4.1 and (4.1) yield

$$\mathbb{P}\left[\gcd(\mathbf{k}) \text{ divides } \sum_{i=1}^{n} \mathbf{d}_{i} \text{ and } \left| dn - \sum_{i=1}^{n} \mathbf{d}_{i} \right| \le \sqrt{n} \,|\, \mathcal{M} \right] = \Omega_{n}(1). \tag{4.5}$$

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

$$\mathbb{P}\left[S'+S''=\sum_{i=1}^{n}\boldsymbol{d}_{i} \mid \gcd(\boldsymbol{k}) \text{ divides } \sum_{i=1}^{n}\boldsymbol{d}_{i}, \left| dn-\sum_{i=1}^{n}\boldsymbol{d}_{i} \right| \leq \sqrt{n}, \left|S'-\mathbb{E}[S'\mid\mathcal{M}]\right| \leq \sqrt{n}, \mathcal{M}\right] = \Omega_{n}(n^{-1/2}). \tag{4.6}$$

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

4.2. **The event** \mathscr{S} . The random matrix A_n for Theorem 1.1 is identical in distribution to the random matrix $A_{0,n}$ with $\varepsilon = 0$ conditioned on the event \mathscr{D} and on the event \mathscr{S} that the Tanner graph $G_{0,n}$ does not contain any multiedges. Therefore, Proposition 1.7 is going to be a consequence of Lemma 4.2 and the following statement.

Lemma 4.3. We have
$$\mathbb{P}\left[A_{0,n} \in \mathcal{S} | \mathcal{D}\right] = \Omega_n(1)$$
.

We proceed to prove Lemma 4.3. Recall the event \mathcal{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 $G_{0,n}$ is asymptotically Poisson with a finite mean. This argument has been carried out illustratively for the case of random regular graphs in [46, 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
$$\mathfrak{D} \cap \mathcal{M}$$
 we have $\frac{1}{n} \sum_{i=1}^{n} d_i^2 \to \mathbb{E}[d^2]$, $\frac{1}{n} \sum_{i=1}^{m} k_i^2 \to d\mathbb{E}[k^2]/k$ in probability.

Proof. We will only prove the statement about the k_i ; the same (actually slightly simplified) argument applies to the d_i . Thanks to Bennett's tail bound for the Poisson distribution we may condition on $\{m=m\}$ 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 m = m the variables $Q_i = \sum_{i \in [m]} 1\{k_i = j\}$ have a binomial distribution. Therefore, the Chernoff bound yields

$$\mathbb{P}\left[\left|Q_{i}-dn\mathbb{P}\left[\mathbf{k}=j\right]/k\right| \leq \sqrt{n}\ln n \mid \mathbf{m}=m\right] = 1 - o_{n}(1/n) \qquad \text{for any } j \leq L.$$

Hence, (4.1) and Lemma 4.2 yield

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

Further, let

$$\begin{split} R_h &= \sum_{j \geq 1} \mathbf{1}\{(1+\delta)^{h-1}L < j \leq (1+\delta)^h L \wedge \sqrt{n}/\ln^9 n\} Q_j, \\ \bar{R}_h &= m \sum_{j \geq 1} \mathbf{1}\{(1+\delta)^{h-1}L < j \leq (1+\delta)^h L \wedge \sqrt{n}/\ln^9 n\} \mathbb{P}\left[\pmb{k} = j\right] \end{split}$$

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

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

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

$$\begin{split} \frac{1}{n} \sum_{i=1}^{m} \pmb{k}_{i}^{2} &\leq o_{n}(1) + \frac{d}{k} \mathbb{E} \left[\pmb{k}^{2} \pmb{1} \{ \pmb{k} \leq L \} \right] + \frac{d}{kn} \sum_{h \in \mathcal{H}} (1+\delta)^{2h} L^{2} R_{h} \\ &= o_{n}(1) + \frac{d}{k} \mathbb{E} \left[\pmb{k}^{2} \pmb{1} \{ \pmb{k} \leq L \} \right] + \frac{d}{kn} \sum_{h \in \mathcal{H}} (1+\delta)^{2h+1} L^{2} \bar{R}_{h} \leq \frac{(1+\delta)d}{k} \mathbb{E} [\pmb{k}^{2}] + o_{n}(1), \quad \text{and analogously} \\ \frac{1}{n} \sum_{i=1}^{m} \pmb{k}_{i}^{2} &\geq \frac{(1-\delta)d}{k} \mathbb{E} [\pmb{k}^{2}] + o_{n}(1). \end{split}$$

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 $G_{0,n}$ and let $\ell \ge 1$. There is $\lambda > 0$ such that on

$$\mathcal{M} \cap \mathcal{D} \cap \left\{ \sum_{i=1}^{n} \boldsymbol{d}_{i} = dn + o_{n}(n), \sum_{i=1}^{n} \boldsymbol{d}_{i}^{2} = n\mathbb{E}[\boldsymbol{d}^{2}] + o_{n}(n) \right\} \cap \left\{ \sum_{i=1}^{m} \boldsymbol{k}_{i}^{2} = dn\mathbb{E}[\boldsymbol{k}^{2}]/k + o_{n}(n) \right\} \cap \left\{ \boldsymbol{m} = dn/k + o_{n}(n) \right\}$$

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

Proof. To estimate the ℓ -th factorial moments of Y for $\ell \ge 1$, 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 $u_i \in [m]$, $v_i \in [n]$ and $2 \le w_i \le k_{u_i} \land d_{v_i} \le \sqrt{n}/\log^9 n$ 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

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

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

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

$$\begin{split} \mathbb{E}[Y_{\boldsymbol{w}} \mid (\boldsymbol{d}_{i})_{i \in [n]}, (\boldsymbol{k}_{i})_{i \in \boldsymbol{m}}] &\leq O_{n}(n^{-w}) \prod_{i=1}^{\ell} \left(\sum_{j=1}^{n} \boldsymbol{d}_{j}^{w_{i}} \right) \left(\sum_{j=1}^{m} \boldsymbol{k}_{j}^{w_{i}} \right) \leq O_{n}(n^{-w}) \max_{j \in [n]} \boldsymbol{d}_{j}^{w-2\ell} \max_{j \in [m]} \boldsymbol{k}_{j}^{w-2\ell} \left(\sum_{j=1}^{n} \boldsymbol{d}_{j}^{2} \right)^{\ell} \left(\sum_{j=1}^{m} \boldsymbol{k}_{j}^{2} \right)^{\ell} \\ &\leq O_{n}(n^{2\ell-w}) \max_{i \in [n]} \boldsymbol{d}_{j}^{w-2\ell} \max_{i \in [m]} \boldsymbol{k}_{j}^{w-2\ell} = O_{n}(\ln^{2\ell-w} n); \end{split}$$

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

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

Further, invoking (4.9), we obtain

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

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

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

Finally, on $\left\{\sum_{i=1}^{n} d_{i} = dn + o_{n}(n), \sum_{i=1}^{n} d_{i}^{2} = n\mathbb{E}[d^{2}] + o_{n}(n)\right\} \cap \left\{\sum_{i=1}^{m} k_{i}^{2} = dn\mathbb{E}[k^{2}]/k + o_{n}(n)\right\}$ 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)–(4.13).

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

Proof. Claims 4.4 and 4.5 show together with inclusion/exclusion (e.g., [12, Theorem 1.22]) that w.h.p. on $\mathcal{M} \cap \mathcal{D}$,

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

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 1.7. The proposition is immediate from Lemmas 4.2 and 4.3. \Box

4.3. **Proof of Proposition 2.5.** The random matrix A_n has n columns and $m \sim \text{Po}(dn/k)$ rows, with the column and row degrees drawn from the distributions d and k. By comparison, $A_{\varepsilon,n}$ has slightly fewer, namely $m_{\varepsilon,n} \sim \text{Po}((1-\varepsilon)dn/k)$ rows. One might therefore think that the proof of Proposition 2.5 is straightforward, as it appears that A_n is obtained from $A_{\varepsilon,n}$ by simply adding another random $\text{Po}(\varepsilon dn/k)$ rows. Since adding $O_{\varepsilon,n}(\varepsilon n)$ rows cannot reduce the nullity by more than $O_{\varepsilon,n}(\varepsilon n)$, the bound on $\mathbb{E}[\text{nul}(A_n)] - \mathbb{E}[\text{nul}(A_{\varepsilon,n})]$ appears to be immediate. But there is a catch. Namely, in constructing A_n we condition on the event $\mathcal{D} = \{\sum_{i=1}^n d_i = \sum_{i=1}^m k_i\}$. Thus, $A_{\varepsilon,n}$ does not have the same distribution as the top $\text{Bin}(m, 1-\varepsilon)$ rows of A_n since the conditioning might distort the degree distribution. We need to show that this distortion is insignificant. To this end, recall that $m_{\varepsilon,n} \sim \text{Po}((1-\varepsilon)dn/k)$.

Lemma 4.7. W.h.p. we have

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

Proof. Lemma 1.8 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 $(\mathfrak{A}_t)_{t \leq \sum_{i=1}^{m_{\varepsilon,n}} \boldsymbol{k}_i}$ that reveals the random matching $\Gamma_{\varepsilon,n}$ one edge at a time. Then

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

П

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

Let $A_{n,\mathcal{D}}$ be the conditional version of the random matrix $A_{0,n}$ given \mathcal{D} . Thus, given $\sum_{i=1}^n d_i = \sum_{i=1}^m k_i$, we construct a random Tanner multi-graph with variable degrees d_1, \ldots, d_n and check degrees k_1, \ldots, k_m . Hence, the difference between A_n and $A_{n,\mathcal{D}}$ is merely that in the case of A_n we also condition on the event \mathcal{S} that the Tanner graph is simple.

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

Proof. Let $G_{n,\mathscr{D}}$ and $G_{\varepsilon,n}$ denote the Tanner graphs corresponding to $A_{n,\mathscr{D}}$ and $A_{\varepsilon,n}$, respectively. It suffices to construct a coupling of $G_{n,\mathscr{D}}$ and $G_{\varepsilon,n}$ 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 $A_{\varepsilon,n}$. Parameter $\mathscr{F} = \mathscr{F}(\varepsilon)$ is given. Generate $\mathbf{\theta} \in [\mathscr{F}]$ uniformly at random. Then generate $\mathbf{m} \sim \operatorname{Po}(dn/k)$ and $\mathbf{m}_{\varepsilon,n} = \operatorname{Bin}(\mathbf{m}, 1 - \varepsilon)$ and then check nodes $a_1, \ldots, a_{\mathbf{m}_{\varepsilon,n}}$. Each check node a_i is associated with an integer \mathbf{k}_i which is an independent copy of \mathbf{k} . To distinguish $G_{\varepsilon,n}$ from $G_{n,\mathscr{D}}$, we colour these check nodes red. Add $\mathbf{\theta}$ check nodes $p_1, \ldots, p_{\mathbf{\theta}}$ to both $G_{\varepsilon,n}$ and $G_{n,\mathscr{D}}$.

Next generate n variable nodes where variable node x_i is associated with d_i , which is an independent copy of d. Further, let $r_j = \sum_{h=1}^m \mathbf{1} \{k_h = j\}$ denote the prospective number of checks of $G_{n,\mathcal{D}}$ 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

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

Hence, Lemma 4.2 implies that

$$\mathbb{P}\left[\sum_{j\geq L} \mathbf{r}_{j} > \varepsilon n \mid \mathcal{D} \cap \mathcal{M}\right] \\
+ \mathbb{P}\left[\exists j \leq L : \mathbf{r}_{j} \leq \sum_{i=1}^{m_{\varepsilon,n}} \mathbf{1}\{\mathbf{k}_{i} = j\} \text{ for all } j \leq L \mid \mathcal{D} \cap \mathcal{M}\right] + \mathbb{P}\left[\mathbf{m} > \mathbf{m}_{\varepsilon,n} + 2\varepsilon dn/k \mid \mathcal{D} \cap \mathcal{M}\right] \leq 1/n = o_{n}(1). \quad (4.14)$$

Now condition on the event

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

Uncolour all (red) check nodes a_i with $k_i \le L$. Moreover, for each $j \le L$, generate $r_j - \sum_{i=1}^{m_{\varepsilon,n}} \mathbf{1}\{k_i = j\}$ additional check nodes of degree j and colour them blue. Finally, for each j > L, generate r_j blue check nodes of degree j.

Now $G_{\varepsilon,n}$ is generated by taking a random maximal matching from the clones of all *uncoloured* and *red* check nodes $\{a_i\} \times [k_i]$ (excluding check nodes p_1, \ldots, p_{θ}) 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 $1 \le i \le \boldsymbol{\theta}$. The Tanner graph $G_{n,\mathcal{D}}$ 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 $1 \le i \le \boldsymbol{\theta}$, 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. \square

Proof of Proposition 2.5. Assume that (2.3) 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

$$\limsup_{n\to\infty} \mathbb{E}[\operatorname{nul}(A_{\varepsilon,n})]/n \le C + \delta.$$

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

$$\mathbb{P}\left[\operatorname{nul}(A_{n,\mathcal{D}})/n \le C + O_{\varepsilon,n}(\varepsilon)\right] > 1 - \varepsilon + o_n(1). \tag{4.15}$$

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

$$\mathbb{P}\left[\operatorname{nul}(A_{n,\mathcal{D}})/n \le C + O_{\varepsilon,n}(\varepsilon) \mid \mathcal{S}\right] > 1 - \delta + o_n(1),\tag{4.16}$$

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

5. THE AIZENMAN-SIMS-STARR SCHEME: PROOF OF PROPOSITION 2.6

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 $A_{\varepsilon,n+1}$ and $A_{\varepsilon,n}$ via Proposition 2.3 and Lemma 2.4. This requires a coupling of the random variables $\text{nul}(A_{\varepsilon,n+1})$ and $\text{nul}(A_{\varepsilon,n})$.

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

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

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

$$\mathbf{M}_{j}^{+} = \mathbf{M}_{j} + \Delta_{j}, \qquad \mathbf{m}_{\varepsilon,n} = \sum_{j\geq 1} \mathbf{M}_{j}, \qquad \mathbf{m}_{\varepsilon,n}^{+} = \sum_{j\geq 1} \mathbf{M}_{j}^{+}.$$
 (5.2)

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

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 $\Gamma_{n,\boldsymbol{M}}$ of the complete bipartite graph with vertex classes

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

Each matching edge $(x_h, s, a_{i,j}, t) \in \Gamma_{n,M}$ 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 $i \in [\theta]$.

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

$$(A_{n,M})_{p_i,x_h} = \mathbf{1} \left\{ i = j \right\}$$
 $(i \in [\theta], h \in [n]),$
$$(A_{n,M})_{a_{i,j},x_h} = \chi_{\zeta_{i,j},\xi_h} \sum_{s=1}^{i} \sum_{t=1}^{d_h} \mathbf{1} \left\{ \{(x_h, t), (a_{i,j}, s)\} \in \Gamma_{n,M} \right\}$$
 $(i \ge 1, j \in [M_i], h \in [n]).$

The Tanner graph G_{n+1,M^+} and its associated random matrix A_{n+1,M^+} are defined analogously.

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

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

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

$$\boldsymbol{M}_{i}^{-} = (\boldsymbol{M}_{i} - \boldsymbol{\gamma}_{i}) \vee 0. \tag{5.3}$$

Consider the random Tanner graph $G' = G_{n,M^-}$ induced by a random maximal matching Γ' of the complete bipartite graph with vertex classes

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

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

Now, obtain the Tanner graph G'' from G' by adding new check nodes $a''_{i,j}$ with target degree i for each $i \ge 1$, $j \in [M_i - M_i^-]$. The new checks are joined by a random maximal matching Γ'' of the complete bipartite graph with vertex classes

$$\mathscr{C}$$
 and $\bigcup_{i\geq 1}\bigcup_{j\in [M_i-M_i^-]}\{a_{i,j}''\}\times [i],$

i.e., for each matching edge we insert a corresponding variable-check edge.

Analogously, obtain G''' by adding one variable node x_{n+1} as well as check nodes $a_{i,j}'', i \ge 1, j \in [\gamma_i]$ and $b_{i,j}'''$, $i \ge 1, j \in [M_i^+ - M_i^- - \gamma_i]$ to G'. The new checks are connected to G' via a random maximal matching Γ''' of the complete bipartite graph with vertex classes

$$\mathscr{C} \qquad \text{and} \qquad \bigcup_{i \geq 1} \Biggl(\bigcup_{j \in [\gamma_i]} \{a_{i,j}'''\} \times [i-1] \cup \bigcup_{j \in [M_i^+ - M_i^- - \gamma_i]} \{b_{i,j}'''\} \times [i] \Biggr).$$

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 A', A'', A''' whose non-zero entries represent the edges of G', G'', G'''. Recalling that $(\zeta_{i,j}, \xi_i)_{i,j\geq 1}$ are uniform on [0,1] and independent of everything else, we additionally introduce independent random variables $(\zeta'_{i,j}, \zeta''_{i,j})_{i,j\geq 1}$, also uniform on [0,1]. With the rows and columns indexed by check and variable nodes, respectively, we define

$$\begin{split} A'_{p_i,j} &= A''_{p_i,j} = A''_{p_i,j} = \mathbf{1} \left\{ i = j \right\} \\ A'_{a_{i,j},x_h} &= A''_{a_{i,j},x_h} = A''_{a_{i,j},x_h} = \chi_{\zeta_{i,j},\xi_h} \sum_{s=1}^{i} \sum_{t=1}^{d_h} \mathbf{1} \left\{ \{(x_h,t),(a_{i,j},s)\} \in \Gamma' \right\} \\ A''_{a''_{i,j},x_h} &= \chi_{\zeta'_{i,j},\xi_h} \sum_{s=1}^{i} \sum_{t=1}^{d_h} \mathbf{1} \left\{ \{(x_h,t),(a''_{i,j},s) \in \Gamma''\} \right\} \\ A'''_{a'''_{i,j},x_h} &= \chi_{\zeta'_{i,j},\xi_h} \sum_{s=1}^{i-1} \sum_{t=1}^{d_h} \mathbf{1} \left\{ \{(x_h,t),(a'''_{i,j},s) \in \Gamma'''\} \right\} \\ A'''_{b''_{i,j},x_h} &= \chi_{\zeta''_{i,j},\xi_h} \sum_{s=1}^{i} \sum_{t=1}^{d_h} \mathbf{1} \left\{ \{(x_h,t),(a'''_{i,j},s) \in \Gamma'''\} \right\} \\ A'''_{b''_{i,j},x_h} &= \chi_{\zeta''_{i,j},\xi_h} \sum_{s=1}^{i} \sum_{t=1}^{d_h} \mathbf{1} \left\{ \{(x_h,t),(b'''_{i,j},s) \in \Gamma'''\} \right\} \\ (i \geq 1, j \in [M_i^+ - M_i^- - \gamma_i], h \in [n]). \end{split}$$

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

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

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

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

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

Proof of Proposition 2.6. The proposition is an immediate consequence of Lemmas 5.1-5.4.

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 d, k, it is not at all difficult to fathom where the right hand side expressions come from. They actually arise naturally from Lemma 2.4 and the scarcity of short proper relations supplied by Proposition 2.3. Indeed, we can write the matrices A'', A''' in the form

$$\mathbf{A}^{\prime\prime} = \begin{pmatrix} \mathbf{A}^{\prime} \\ \mathbf{B} \end{pmatrix}, \qquad \mathbf{A}^{\prime\prime\prime} = \begin{pmatrix} \mathbf{A}^{\prime} & 0 \\ \mathbf{B}^{\prime} & \mathbf{C}^{\prime} \end{pmatrix}$$
 (5.4)

with B, (B' C') representing the new rows and, in the case of A''', the additional column. To calculate $\mathbb{E}[\operatorname{nul}(A''[\theta]) - \operatorname{nul}(A'[\theta])]$ we basically need to assess the impact of adding a few more checks $a''_{i,j}$ to the Tanner graph G' of A'. The new checks connect to randomly chosen cavities of A'. Let k_1, \ldots, k_L denote the degrees of the new checks. Since the distribution k of the check degrees has a finite second moment, the total degree $k_1 + \cdots + k_L$ is bounded w.h.p. The random matrix B therefore encodes the non-zero entries corresponding to the edges that connect the $a'_{i,j}$ with the cavities of A' where the new checks attach. Furthermore, the construction of A' ensures that w.h.p. 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 2.3, applied with $\mathcal{T} = \mathcal{T}(\varepsilon)$ large enough, implies that the probability that the set I of non-zero columns of B forms a proper relation is $o_{\varepsilon}(1)$. Consequently, Lemma 2.4 yields

$$\mathbb{E}[\operatorname{nul}(\mathbf{A}'') - \operatorname{nul}(\mathbf{A}')] = -\mathbb{E}[\operatorname{rk}(\mathbf{B}_*)] + o_n(1), \tag{5.5}$$

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

$$\mathbb{E}[\operatorname{rk}(\boldsymbol{B}_*) \mid \boldsymbol{\alpha}, k_1, \dots, k_L] = \sum_{i=1}^{L} \left(1 - \boldsymbol{\alpha}^{k_i}\right) + o_{\varepsilon, n}(1).$$
(5.6)

Substituting (5.6) into (5.5) and the correct distribution of $k_1, ..., k_L$ supplied by the coupling into (5.6), we obtain the expression displayed in Lemma 5.4. 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 $V_n = \{x_1, ..., x_n\}$ of variables induced by choosing a cavity uniformly at random, i.e.,

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

in the (unlikely) event that $\mathscr{C} = \emptyset$, we use the convention $P(x_1) = 1$. Let $x_1, x_2, ... \in V_n$ be independent samples drawn from P. The following lemma shows that $|\mathscr{C}|$ is linear in n w.h.p.

Lemma 5.5. *W.h.p.* we have $|\mathscr{C}| \ge \varepsilon dn/2$.

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

$$\mathbb{P}\left[\left|\sum_{j\geq 1} j\mathbf{M}_j - (1-\varepsilon)dn\right| \leq \sqrt{n}\log n\right] = 1 - o_n(1). \tag{5.7}$$

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

$$\mathbb{P}\left[\left|\sum_{i=1}^{n} \boldsymbol{d}_{i} - dn\right| \le \sqrt{n} \log n\right] = 1 - o_{n}(1). \tag{5.8}$$

Since $|\mathscr{C}| \ge \sum_{i=1}^{n} d_i - \sum_{j\ge 1} jM_j$ by construction, the assertion follows from (5.7) and (5.8).

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

$$\mathscr{E} = \{ \mathbb{P} \left[\mathbf{x}_1, \dots, \mathbf{x}_{\ell_*} \text{ form a proper relation of } \mathbf{A}' \mid \mathbf{A}' \right] < \delta_* \}. \tag{5.9}$$

The following simple lemma is an application of Proposition 2.3.

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

Proof. Lemma 5.5 provides that $|\mathscr{C}| \ge \varepsilon n/2$ w.h.p. 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 $\mathscr{A} = \mathscr{L} \cap \{|\mathscr{C}| \ge \varepsilon n/2\}$.

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

$$\mathbb{P}\left[\left(\boldsymbol{x}_{1},\ldots,\boldsymbol{x}_{\ell_{*}}\right)\in\mathcal{W}\mid\boldsymbol{A}'\right]\leq\left(2L/\varepsilon\right)^{\ell_{*}}\mathbb{P}\left[\left(\hat{\boldsymbol{x}}_{1},\ldots,\hat{\boldsymbol{x}}_{\ell_{*}}\right)\in\mathcal{W}\mid\boldsymbol{A}'\right]+\delta_{*}^{2}/8.\tag{5.10}$$

Applying (5.10) to the set W of proper relations and invoking Proposition 2.3 completes the proof.

Further, consider the event

$$\mathcal{E}' = \left\{ |\mathcal{C}| \ge \varepsilon dn/2 \wedge \max_{i \le n} d_i \le n^{1/2} \right\}. \tag{5.11}$$

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

Proof. This follows from the choice of the parameters in (5.1), Lemma 1.8 and Lemma 5.5.

To prove Lemmas 5.3 and 5.4 we need an explicit description of the vector γ that captures the degrees of the checks adjacent to the new variable node x_{n+1} . Since γ is defined in terms of the the 'big' Tanner graph G_{n+1,M^+} , γ 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 γ can be expressed easily in terms of the sequence $(\hat{k}_i)_{i\geq 1}$ of independent copies of \hat{k} from (1.10). Indeed, let

$$\hat{\boldsymbol{\gamma}}_j = \sum_{i=1}^{d_{n+2}} \mathbf{1}\{\hat{\boldsymbol{k}}_i = j\}$$
 and
$$\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\gamma}}_j)_{j \ge 1}.$$

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

$$\hat{\Delta}_{j} \sim \text{Po}\left((1-\varepsilon)\mathbb{P}\left[\mathbf{k}=j\right]d/k\right). \tag{5.12}$$

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

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

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

Proof. We begin by studying the unconditional distributions of γ and Δ . Let $\zeta = (\sum_{i \ge 1} i M_i^+)/(\sum_{i=1}^{n+1} d_i)$. Proceeding as in the proof of Lemma 5.5, we conclude that $\mathbb{P}[1-2\varepsilon \le \zeta \le 1-\varepsilon/2] = 1-o_n(1)$. Further, given $1-2\varepsilon \le \zeta \le 1-\varepsilon/2$ we can think of G_{n+1,M^+} as being generated by the following experiment.

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

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

(iii) Obtain G^* with variable nodes $x_1, ..., x_{n+1}$ and check nodes $a_{i,j}, i \ge 1, j \in [M_i^+]$ by inserting an edge between x_h and $a_{i,j}$ for any edge of Γ^* 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 $\mathcal{C} = \mathbf{C}$ of cavities and in the next two steps we connect the non-cavities randomly.

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

$$d_{\text{TV}}\left(\boldsymbol{d}_{n+1}^{\star} \mid \{\boldsymbol{d}_{n+1} \leq \varepsilon^{-1/2}\}, \boldsymbol{d}\right) = O_{\varepsilon,n}(\varepsilon^{1/2}). \tag{5.13}$$

Regarding the degrees of the checks adjacent to x_{n+1} , by the principle of deferred decisions we can construct Γ^* 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 $\boldsymbol{M}_i^+ = (1-\varepsilon)\mathbb{P}\left[\boldsymbol{k}=i\right]dn/k + o_n(n)$ for all $i\leq L$ and $\sum_{i\geq 1}i\boldsymbol{M}_i^+ = (1-\varepsilon)dn + o_n(n)$ w.h.p. 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 $\boldsymbol{M}_i^+ = (1-\varepsilon)\mathbb{P}\left[\boldsymbol{k}=i\right]dn/k + o_n(n)$ for all $i\leq L$ and $\sum_{i\geq 1}i\boldsymbol{M}_i^+ = (1-\varepsilon)dn + o_n(n)$ we have

$$d_{\text{TV}}(\boldsymbol{\gamma}, \hat{\boldsymbol{\gamma}}) = O_{\varepsilon,n}(\varepsilon^{1/2}). \tag{5.14}$$

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

To complete the proof we are going to argue that M^-, d_1, \dots, d_n and γ, Δ are asymptotically independent. Arguing along the lines of the previous paragraph, we find that for large $L = L(\varepsilon) > 0$ the event

$$\mathcal{K} = \left\{ \sum_{i>1} i(\Delta_i + \gamma_i) \le L \right\}$$

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

$$\mathcal{L} = \{ \mathbb{P}[\mathcal{K} \mid \mathbf{M}^-, \mathbf{d}_1, \dots, \mathbf{d}_n] \ge 1 - \exp(-1/\varepsilon) \}$$

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

$$\mathcal{M} = \left\{ \forall i \leq L : |\boldsymbol{M}_{i}^{-} - \mathbb{E}[\boldsymbol{M}_{i}]| \leq \sqrt{n} \ln n \right\} \cap \left\{ \sum_{i=1}^{n} \boldsymbol{d}_{i} = (1 - \varepsilon) dn + o_{n}(n) \right\} \cap \left\{ \sum_{i \geq 1} i \boldsymbol{M}_{i}^{-} = (1 - \varepsilon) dn + o_{n}(n) \right\}$$

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

$$\mathbb{P}[\mathcal{X}] \ge \exp(-1/\varepsilon^2), \qquad \mathbb{P}[\mathcal{L}] \ge 1 - \exp(-1/\varepsilon), \qquad \mathbb{P}[\mathcal{M} \mid \mathcal{X}] = 1 - o_n(1). \tag{5.15}$$

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

$$\mathbb{P}\left[\boldsymbol{\gamma} = \boldsymbol{\gamma}, \boldsymbol{\Delta} = \Delta \mid \boldsymbol{M}^{-} = \boldsymbol{M}^{-}, \forall i \in [n] : \boldsymbol{d}_{i} = d_{i}\right] \sim \mathbb{P}\left[\boldsymbol{\gamma} = \boldsymbol{\gamma}\right] \mathbb{P}\left[\boldsymbol{\Delta} = \Delta\right]. \tag{5.16}$$

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

$$\mathbb{P}\left[\mathbf{M}^{-} = \mathbf{M}^{-}, \forall i \leq n : \mathbf{d}_{i} = d_{i} \mid \boldsymbol{\gamma} = \boldsymbol{\gamma}, \boldsymbol{\Delta} = \Delta\right] = \mathbb{P}\left[\mathbf{M} = \mathbf{M}^{-} + \boldsymbol{\gamma}, \forall i \leq n : \mathbf{d}_{i} = d_{i} \mid \boldsymbol{\gamma} = \boldsymbol{\gamma}, \boldsymbol{\Delta} = \Delta\right]$$

$$= \frac{\mathbb{P}\left[\boldsymbol{\gamma} = \boldsymbol{\gamma}, \boldsymbol{\Delta} = \Delta \mid \mathbf{M} = \mathbf{M}^{-} + \boldsymbol{\gamma}, \forall i \leq n : \mathbf{d}_{i} = d_{i}\right]}{\mathbb{P}\left[\boldsymbol{\gamma} = \boldsymbol{\gamma}, \boldsymbol{\Delta} = \Delta\right]} \cdot \mathbb{P}\left[\mathbf{M} = \mathbf{M}^{-} + \boldsymbol{\gamma}\right] \cdot \prod_{i=1}^{n} \mathbb{P}\left[\mathbf{d}_{i} = d_{i}\right]$$

$$= (1 + o_{n}(1)) \frac{\mathbb{P}\left[\boldsymbol{\gamma} = \boldsymbol{\gamma} \mid \mathbf{M} = \mathbf{M}^{-} + \boldsymbol{\gamma}, \forall i \leq n : \mathbf{d}_{i} = d_{i}, \boldsymbol{\Delta} = \Delta\right]}{\mathbb{P}\left[\boldsymbol{\gamma} = \boldsymbol{\gamma}\right]} \cdot \mathbb{P}\left[\mathbf{M} = \mathbf{M}^{-}\right] \cdot \prod_{i=1}^{n} \mathbb{P}\left[\mathbf{d}_{i} = d_{i}\right].$$
(5.17)

Finally, given $M = M^- + \gamma$ and $\Delta = \Delta$ we have $M_i^+ = (1 - \varepsilon) \mathbb{P}[k = i] dn/k + o_n(n)$ for all $i \le L$ and $\sum_{i \ge 1} i M_i^+ = (1 - \varepsilon) dn + o_n(n)$. Therefore, by the principle of deferred decisions, once we condition on a likely outcomes M^- of M^- , d_1, \ldots, d_n and of Δ , the conditional probability of obtaining $\gamma = \gamma$ is close to the unconditional probability:

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

Hence, (5.16) follows from (5.15) and (5.17).

Finally, to complete the proof we combine (5.15) and (5.16) to conclude that with probability $1-\exp(-\Omega_{\varepsilon,n}(1/\varepsilon))$,

$$\mathbb{P}\left[\boldsymbol{\gamma} = \boldsymbol{\gamma}, \boldsymbol{\Delta} = \Delta \mid \boldsymbol{\Sigma}'\right] = \mathbb{P}\left[\boldsymbol{\gamma} = \boldsymbol{\gamma}, \boldsymbol{\Delta} = \Delta \mid \boldsymbol{M}^{-}, \boldsymbol{d}_{1}, \dots, \boldsymbol{d}_{n}\right] = (1 + o_{n}(1))\mathbb{P}\left[\boldsymbol{\gamma} = \boldsymbol{\gamma}\right]\mathbb{P}\left[\boldsymbol{\Delta} = \Delta\right]. \tag{5.18}$$

The assertion follows from (5.14) and (5.18).

5.3. Proof of Lemma 5.3. The proof comprises several steps, each relatively simple individually. Let

$$X = \sum_{i \ge 1} \Delta_i, \qquad Y = \sum_{i \ge 1} i \Delta_i, \qquad Y' = \sum_{i \ge 1} i \gamma_i.$$

Then the total number of new non-zero entries upon going from A' to A''' is bounded by Y + Y'. Let

$$\mathcal{E}'' = \{X \vee Y \vee Y' \leq 1/\varepsilon\}.$$

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,

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

Thus, the assertion follows from Markov's inequality.

Going from G' to G''' we add checks $a_{i,j}'''$, $i \ge 1$, $j \in [\gamma_i]$ and $b_{i,j}'''$, $i \ge 1$, $j \in [M_i^+ - M_i^- - \gamma_i]$. Let

$$\mathcal{X} = \left(\bigcup_{i \ge 1} \bigcup_{j=1}^{\gamma_i} \partial a_{i,j}''' \setminus \{x_{n+1}\}\right) \cup \left(\bigcup_{i \ge 1} \bigcup_{j \in [M_i^+ - M_i^- - \gamma_i]} \partial b_{i,j}'''\right)$$

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

$$\mathscr{E}''' = \left\{ |\mathscr{X}| = Y + \sum_{i \ge 1} (i - 1) \gamma_i \right\}$$

be the event that the variables of G' where the new checks attach are all distinct.

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

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 neighbours of the new checks. Thus, the probability of hitting the same variable twice is $o_n(1)$.

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

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

$$\mathbb{E}\left[\left|\operatorname{nul}(\boldsymbol{A}^{\prime\prime\prime}) - \operatorname{nul}(\boldsymbol{A}^{\prime})\right| (1 - \mathbf{1}\mathcal{E}^{\prime\prime})\right] \le \mathbb{E}\left[\left(X + \boldsymbol{d}_{n+1} + 1\right)^{2}\right]^{1/2} \left(1 - \mathbb{P}\left[\mathcal{E}^{\prime\prime}\right]\right)^{1/2} = o_{\varepsilon,n}(1). \tag{5.19}$$

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

$$\mathbb{E}\left[\left|\operatorname{nul}(A''') - \operatorname{nul}(A')\right| \mathbf{1}\mathscr{E}'' \setminus \mathscr{E}\right] \le O_{\varepsilon,n}(\varepsilon^{-1}) \exp(-1/\varepsilon^4) = o_{\varepsilon,n}(1), \tag{5.20}$$

$$\mathbb{E}\left[\left|\operatorname{nul}(A''') - \operatorname{nul}(A')\right| \mathbf{1}\mathscr{E}'' \setminus \mathscr{E}'\right], \mathbb{E}\left[\left|\operatorname{nul}(A''') - \operatorname{nul}(A')\right| \mathbf{1}\mathscr{E}'' \cap \mathscr{E}' \setminus \mathscr{E}'''\right] = o_n(1). \tag{5.21}$$

The assertion follows from (5.19)–(5.21).

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

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

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

Proof. Let

$$\mathcal{A} = \left\{ a_{i,j}^{\prime\prime\prime} : i \ge 1, \ j \in [\boldsymbol{\gamma}_i] \right\}$$

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

$$\mathscr{B} = \left\{ b_{i,j}^{\prime\prime\prime} : i \ge 1, \ j \in [\boldsymbol{M}_i^+ - \boldsymbol{M}_i^- - \boldsymbol{\gamma}_i] \right\}$$

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 $\mathcal{A} \cup \mathcal{B}$ and whose columns are indexed by $V_n = \{x_1, \dots, x_n\}$ such that for each $a \in \mathcal{A} \cup \mathcal{B}'$ 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 \mathcal{E}''' occurs, then each column of \boldsymbol{B} contains at most one non-zero entry and each row contains at least one non-zero entry. In effect, \boldsymbol{B} has full rank, i.e.,

$$\operatorname{rk}(\boldsymbol{B}) = |\mathcal{A} \cup \mathcal{B}| = \sum_{i>1} \boldsymbol{M}_i^+ - \boldsymbol{M}_i^-.$$

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

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

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

Moreover, on \mathscr{E}' the set \mathscr{X}''' of non-zero columns of 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 2.4 implies that on $\mathscr{E} \cap \mathscr{E}' \cap \mathscr{E}''$,

$$\mathbb{E}\left[\left(\operatorname{nul}(\mathbf{A}''') - \operatorname{nul}(\mathbf{A}')\right) \mathbf{1}\mathscr{E}''' \mid \Sigma''\right] = 1 - \mathbb{E}\left[\operatorname{rk}(\mathbf{B}_* \mathbf{C}) \mid \Sigma''\right] + o_{\varepsilon,n}(1). \tag{5.22}$$

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

$$\mathbb{E}\left[\operatorname{rk}(\boldsymbol{Q}_{\mathscr{B}}) \mid \boldsymbol{\Sigma}''\right] = o_n(1) + \sum_{i \ge 1} \left(1 - \boldsymbol{\alpha}^i\right) (\boldsymbol{M}_i^+ - \boldsymbol{M}_i^- - \boldsymbol{\gamma}_i), \tag{5.23}$$

$$\mathbb{E}\left[\operatorname{rk}(\boldsymbol{Q}_{\mathscr{A}}) \mid \boldsymbol{\Sigma}''\right] = o_n(1) + \sum_{i \ge 1} \left(1 - \boldsymbol{\alpha}^{i-1}\right) \boldsymbol{\gamma}_i + 1 - \prod_{i \ge 1} \left(1 - \boldsymbol{\alpha}^{i-1}\right)^{\boldsymbol{\gamma}_i},\tag{5.24}$$

where, as we recall, α is the probability that a cavity chosen from $p(\cdot)$ is frozen. Indeed, the probability that a \mathcal{B} -row of \mathbf{B} that contains precisely i non-zero entries gets zeroed out completely in \mathbf{B}_* equals $\alpha^i + o_n(1)$ and there are $\mathbf{M}_i^+ - \mathbf{M}_i^- - \mathbf{\gamma}_i$ such rows; hence (5.23).

Similarly, the probability that an \mathscr{A} -row of B with i-1 non-zero entries gets zeroed out completely in B_* equals $\alpha^{i-1}+o_n(1)$ and there are γ_i such rows. Hence, the expected rank of the \mathscr{A} -rows of B_* equals $\sum_{i\geq 1}\left(1-\alpha^{i-1}\right)\gamma_i+o_n(1)$, which is the first summand in (5.24). Moreover, the presence of the C-column adds one to the rank of $Q_{\mathscr{A}}$ unless not a single one of the \mathscr{A} -rows of B gets zero out, which occurs with probability $\prod_{i\geq 1}\left(1-\alpha^{i-1}\right)^{\gamma_i}+o_n(1)$. Hence, we obtain (5.24). Finally, the assertion follows from (5.22)–(5.24).

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

$$\mathbb{E}\left|\mathbb{E}\left[\operatorname{nul}(\boldsymbol{A}^{\prime\prime\prime}) - \operatorname{nul}(\boldsymbol{A}^{\prime}) \mid \boldsymbol{\Sigma}^{\prime\prime}\right] - \left(\prod_{i \geq 1} (1 - \boldsymbol{\alpha}^{i-1})^{\boldsymbol{\gamma}_i} - \sum_{i \geq 1} (1 - \boldsymbol{\alpha}^{i-1}) \boldsymbol{\gamma}_i - \sum_{i \geq 1} (1 - \boldsymbol{\alpha}^i) (\boldsymbol{M}_i^+ - \boldsymbol{M}_i^- - \boldsymbol{\gamma}_i)\right) \mathbf{1}\mathfrak{E}\right| = o_{\varepsilon,n}(1). \quad (5.25)$$

Since on $\mathfrak E$ all degrees i with $M_i^+ - M_i^- - \gamma_i > 0$ are bounded and Chebyshev's inequality shows that $M_i \sim \mathbb E[M_i] = \Omega_n(n)$ for any fixed i w.h.p., (5.3) yields $M_i^- = M_i - \gamma_i$ w.h.p. Hence, (5.25) turns into

$$\mathbb{E}\left|\mathbb{E}\left[\operatorname{nul}(A''') - \operatorname{nul}(A') \mid \Sigma''\right] - \left(\prod_{i \ge 1} (1 - \boldsymbol{\alpha}^{i-1})^{\boldsymbol{\gamma}_i} - \sum_{i \ge 1} (1 - \boldsymbol{\alpha}^{i-1}) \boldsymbol{\gamma}_i - \sum_{i \ge 1} (1 - \boldsymbol{\alpha}^i) \boldsymbol{\Delta}_i\right) \mathbf{1} \mathcal{E}''\right| = o_{\varepsilon,n}(1). \tag{5.26}$$

Further, since $\sum_{i\geq 1} \gamma_i \leq d_{n+1}$ and $\mathbb{E}[d_{n+1}] = O_{\varepsilon,n}(1)$, we obtain

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

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

$$=\mathbb{E}\left[\left(\prod_{i\geq 1}(1-\boldsymbol{\alpha}^{i-1})^{\boldsymbol{\gamma}_i}-\sum_{i\geq 1}(1-\boldsymbol{\alpha}^{i-1})\boldsymbol{\gamma}_i\right)\mathbf{1}\left\{\sum_{i\geq 1}\boldsymbol{\gamma}_i\leq \varepsilon^{-1/4}\right\}\right]+o_{\varepsilon,n}(1) \qquad \text{[by Lemmas 5.6-5.7/Claims 5.9-5.10]}$$

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

$$=\mathbb{E}\left[(1-\boldsymbol{\alpha}^{\hat{k}-1})^{\hat{\boldsymbol{d}}}-\boldsymbol{d}-\boldsymbol{d}\boldsymbol{\alpha}^{\hat{k}-1}\right]+o_{\varepsilon,n}(1) \qquad \text{[by the def. of }\hat{\boldsymbol{\gamma}}]$$

$$=\mathbb{E}\left[D(1-K'(\boldsymbol{\alpha})/k)-d-\frac{d}{k}K'(\boldsymbol{\alpha})\right]+o_{\varepsilon,n}(1) \qquad \text{[by (1.10)]}. \qquad (5.27)$$

Similarly, Claim 5.9, Lemma 5.8 and the construction (5.12) of $\hat{\Delta}$ yield

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

$$= o_{\varepsilon,n}(1) + (1-\varepsilon)\frac{d}{k}\sum_{i\geq 1}\mathbb{P}\left[\boldsymbol{k}=i\right]\mathbb{E}\left[1-\boldsymbol{\alpha}^{i}\right] = o_{\varepsilon,n}(1) + \frac{d}{k} - \frac{d}{k}\mathbb{E}\left[K(\boldsymbol{\alpha})\right].$$
(5.28)

Finally, the assertion follows from (5.26), (5.27) and (5.28).

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 A'' from A'. As before we consider the events $\mathscr{E}, \mathscr{E}'$ from (5.9) and (5.11) Moreover, recalling that the total number of new non-zero entries when going from A' to A'' is bounded by d_{n+1} , we introduce $\mathscr{E}'' = \{d_{n+1} \le 1/\epsilon\}$.

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

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

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

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

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 $\mathbb{P}\left[\mathscr{E}''' \mid \mathscr{E}' \cap \mathscr{E}''\right] = 1 - o_n(1)$.

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 $\mathbb{E}\left[\left|\operatorname{nul}(A'') - \operatorname{nul}(A')\right| (1 - 1\mathcal{E} \cap \mathcal{E}' \cap \mathcal{E}'' \cap \mathcal{E}''')\right] = o_{\varepsilon,n}(1)$.

Proof. We have $|\operatorname{nul}(A'') - \operatorname{nul}(A')| \le d_{n+1}$ as we add at most d_{n+1} rows. Because $\mathbb{E}[d_{n+1}] = O_{\varepsilon,n}(1)$, Claim (5.13) and the Cauchy-Schwarz inequality yield

$$\mathbb{E}\left[\left|\operatorname{nul}(A'') - \operatorname{nul}(A')\right| (1 - 1\mathscr{E}'')\right] \le \mathbb{E}\left[d_{n+1}^2\right]^{1/2} (1 - \mathbb{P}\left[\mathscr{E}\right])^{1/2} = o_{\varepsilon,n}(1). \tag{5.29}$$

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

$$\mathbb{E}\left[\left|\operatorname{nul}(A'') - \operatorname{nul}(A')\right| \mathbf{1}\mathcal{E}'' \setminus \mathcal{E}\right], \mathbb{E}\left[\left|\operatorname{nul}(A'') - \operatorname{nul}(A')\right| \mathbf{1}\mathcal{E}'' \setminus \mathcal{E}'\right], \mathbb{E}\left[\left|\operatorname{nul}(A'') - \operatorname{nul}(A')\right| \mathbf{1}\mathcal{E}'' \setminus \mathcal{E}'''\right] = o_{\varepsilon,n}(1). \tag{5.30}$$

The assertion follows from (5.29) and (5.30).

The matrix A'' results from A' by adding checks $a''_{i,j}$, $i \ge 1$, $j \in [M_i - M_i^-]$ that are connected to random cavities of A'. Moreover, as before let $\Sigma'' \supset \Sigma'$ be the σ -algebra generated by G', A', M_- , $(d_i)_{i \in [n+1]}$, γ , M and Δ . 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 $\mathbb{E} \left[(\operatorname{nul}(A'') - \operatorname{nul}(A')) \mathbf{1} \mathscr{E}''' \mid \Sigma'' \right] = o_{\varepsilon,n}(1) - \sum_{i>1} (1 - \boldsymbol{\alpha}^i) (M_i - M_i^-).$

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

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

$$\operatorname{nul} \mathbf{A}'' = \operatorname{nul} \begin{pmatrix} \mathbf{A}' \\ \mathbf{B} \end{pmatrix}. \tag{5.31}$$

Moreover, on $\mathcal{E}' \cap \mathcal{E}''$ the set \mathcal{X}'' of non-zero columns of \mathbf{B} has size at most $|\mathcal{X}''| \leq \mathbf{d}_{n+1} \leq 1/\varepsilon$, while there are at least $\varepsilon dn/2$ cavities. Hence, on $\mathcal{E} \cap \mathcal{E}' \cap \mathcal{E}'' \cap \mathcal{E}'''$ the probability that \mathcal{X}'' forms a proper relation is bounded by $\exp(-1/\varepsilon^4)$. Therefore, Lemma 2.4 implies that

$$\mathbb{E}\left[\left(\operatorname{nul}(\mathbf{A}'') - \operatorname{nul}(\mathbf{A}')\right) \mathbf{1}\mathscr{E}''' \mid \Sigma''\right] = o_{\varepsilon,n}(1) - \mathbb{E}\left[\operatorname{rk}(\mathbf{B}_*) \mid \Sigma''\right]. \tag{5.32}$$

Further, since an α -fraction of cavities are frozen, a row of B with i non-zero entries gets zeroed out completely in B_* with probability $\alpha^i + o_n(1)$. Consequently,

$$\mathbb{E}\left[\operatorname{rk}(\boldsymbol{B}_*) \mid \boldsymbol{\Sigma}''\right] = o_{\varepsilon,n}(1) + \sum_{i \ge 1} \left(1 - \boldsymbol{\alpha}^i\right) (\boldsymbol{M}_i - \boldsymbol{M}_i^-). \tag{5.33}$$

Finally, the assertion follows from (5.32) and (5.33).

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

$$\mathbb{E}\left|\mathbb{E}[\operatorname{nul}(A'') - \operatorname{nul}(A') \mid \Sigma''] + \left(\sum_{i \ge 1} (1 - \boldsymbol{\alpha}^i)(\boldsymbol{M}_i - \boldsymbol{M}_i^-)\right) \mathbf{1}\mathfrak{E}\right| = o_{\varepsilon,n}(1). \tag{5.34}$$

Since on $\mathfrak E$ all degrees i with $M_i^+ - M_i^- > 0$ are bounded w.h.p. and $M_i^- = \Omega_n(n)$ w.h.p., we conclude that $M_i - M_i^- = \gamma_i$ for all $i \ge 1$ w.h.p. Hence, (5.34) turns into

$$\mathbb{E}\left|\mathbb{E}[\mathrm{nul}(A'') - \mathrm{nul}(A') \mid \Sigma''] + \left(\sum_{i \ge 1} (1 - \boldsymbol{\alpha}^i) \boldsymbol{\gamma}_i\right) \mathbf{1}\mathfrak{E}\right| = o_{\varepsilon, n}(1). \tag{5.35}$$

Further, because $\sum_{i\geq 1} \gamma_i \leq d_{n+1}$ and $\mathbb{E}[d_{n+1}] = O_{\varepsilon,n}(1)$,

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

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

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

The assertion follows from (5.35) and (5.36).

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 $\mathbb{E}[\operatorname{nul}(A'')] = \mathbb{E}[\operatorname{nul}(A_{n,M})] + o_n(1)$.

Proof. The choice of the random variables in (5.1) and Lemma 1.8 ensure that the event $\mathscr{E} = \{\sum_{i \geq 1} i M_i \leq dn/k\}$ has probability $1 - o_n(1/n)$. Further, given \mathscr{E} the random variables $\operatorname{nul}(A'')$ and $\operatorname{nul}(A_{n,M})$ 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 $\operatorname{nul}(A''')$ and $\operatorname{nul}(A_{n+1,M^+})$ we consider the event

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

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

Proof. This follows from the definition (5.2) of the random variables M_i^+ and Lemma 1.8.

Further, consider the event

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

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

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

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

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

Proof. Given $W \cap \mathcal{E}^+$ 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 $\operatorname{nul}(A''')$, $\operatorname{nul}(A_{n+1,M^+})$ can be coupled identically on the 'bulk' event $\mathcal{E}^+ \cap \mathcal{U} \cap \mathcal{W}$.

Claim 5.21. Given $\mathcal{E}^+ \cap \mathcal{U} \cap \mathcal{W}$ the random variables $\operatorname{nul}(A''')$ and $\operatorname{nul}(A_{n+1,M^+})$ are identically distributed and thus $\mathbb{E}\left[\left(\operatorname{nul}(A''') - \operatorname{nul}(A_{n+1,M^+})\right) \mathbf{1} \mathcal{U} \cap \mathcal{W} \cap \mathcal{E}^+\right] = 0. \tag{5.37}$

Proof. By construction, on $\mathscr{E}^+ \cap \mathscr{U} \cap \mathscr{W}$ the random matrices A''' and A_{n+1,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 $\mathcal{E}^+ \setminus (\mathcal{U} \cap \mathcal{W})$.

Claim 5.22. There is a coupling of A_{n+1,M^+} and A''' on \mathcal{E}^+ such that $|\operatorname{nul}(A''') - \operatorname{nul}(A_{n+1,M^+})| \le 2\sum_{i\ge 1} i(\Delta_i + \gamma_i)$.

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

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

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

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

$$\mathcal{Q}_1 = \mathcal{E}^+ \cap \left\{ \exists i > \log n : \boldsymbol{\gamma}_i > 0 \right\}, \quad \mathcal{Q}_2 = \mathcal{E}^+ \cap \left\{ \boldsymbol{d}_{n+1} > \log n \right\} \setminus \mathcal{Q}_1, \quad \mathcal{Q}_3 = \mathcal{E}^+ \cap \left\{ \sum_{i \geq 1} i \boldsymbol{\Delta}_i > \ln^3 n \right\} \setminus (\mathcal{Q}_1 \cup \mathcal{Q}_2).$$

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

$$\mathbb{E}\left[\sum_{i\geq 1}i\boldsymbol{\gamma}_{i}\mathbf{1}\mathcal{Q}_{1}\right]\leq O_{n}(1/n)\cdot\left(1+\mathbb{E}\left[\sum_{i=1}^{\boldsymbol{m}_{\epsilon,n}^{+}}\mathbf{1}\{\boldsymbol{k}_{i}\geq\log n\}\boldsymbol{k}_{i}^{2}\boldsymbol{d}_{n+1}\right]\right)=O_{n}(1)\cdot\mathbb{P}\left[\boldsymbol{k}\geq\log n\right]+O_{n}(1/n)=O_{n}(\log^{-2}n).$$
(5.38)

Indeed, the last equality sign follows from the first because $\mathbb{E}[k^2] = O_n(1)$ and the first equality sign follows because $m_{\varepsilon,n}^+$ is independent of d_{n+1} and the k_i . Further, to obtain the first inequality we consider the $m_{\varepsilon,n}^+$ checks one by one. The degree of the ith check is distributed as k_i . We discard it unless $k_i \ge \log n$. But if $k_i \ge \log n$, then the probability that k_i is adjacent to x_{n+1} is bounded by $O_n(k_i d_{n+1}/\sum_{h=1}^{n+1} d_h)$ and $\sum_{h=1}^{n+1} d_h \ge n$. Thus, we obtain (5.38). Further, we observe that (5.38) yields $\mathbb{P}[\mathcal{Q}_1] \le \mathbb{E}\sum_{i\ge 1} i\gamma_i 1\mathcal{Q}_1 = O_n(\log^{-2} n)$. Hence, as $\mathbb{E}\sum_{i\ge 1} i\Delta_i = O_n(1)$ we obtain

$$\mathbb{E}\left[\sum_{i\geq 1} i\Delta_i \mathbf{1} \mathcal{Q}_1\right] \leq \mathbb{P}\left[\mathcal{Q}_1\right] \log n + \mathbb{E}\left[\sum_{i\geq 1} i\Delta_i \mathbf{1} \left\{\sum_{i\geq 1} i\Delta_i \geq \log n\right\}\right] = o_n(1). \tag{5.39}$$

Combining (5.38) and (5.39), we conclude that

$$\mathbb{E}\left[\sum_{i\geq 1}i(\boldsymbol{\Delta}_i+\boldsymbol{\gamma}_i)\mathbf{1}\mathcal{Q}_1\right]=o_n(1). \tag{5.40}$$

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

$$\mathbb{E}\left[\sum_{i\geq 1}i\boldsymbol{\gamma}_{i}\mathbf{1}\mathcal{Q}_{2}\right]\leq O_{n}(\log n)\mathbb{E}\left[\boldsymbol{d}_{n+1}\mathbf{1}\{\boldsymbol{d}_{n+1}>\log n\}\right]=o_{n}(1). \tag{5.41}$$

Moreover, since the Δ_i are independent of d_{n+1} and $\mathbb{E}\sum_{i\geq 1}i\Delta_i=O_n(1)$, we obtain $\mathbb{E}\left[\sum_{i\geq 1}i\Delta_i\mathbf{1}\mathcal{Q}_2\right]=o_n(1)$. Hence, (5.41) yields

$$\mathbb{E}\left[\sum_{i\geq 1}i(\mathbf{\Delta}_i+\boldsymbol{\gamma}_i)\mathbf{1}\mathcal{Q}_2\right]=o_n(1). \tag{5.42}$$

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

$$\mathbb{P}\left[\mathcal{Q}_3\right] \le \mathbb{E}\left[\sum_{i \ge 1} i\Delta_i\right] \ln^{-3} n = O_n(\mathbb{E}[\mathbf{k}^2] \ln^{-3} n) = o_n(\log^{-2} n). \tag{5.43}$$

Moreover, on \mathcal{Q}_3 we have $\sum_{i\geq 1} i \gamma_i \leq \log^2 n$ because $\boldsymbol{d}_{n+1} \leq \log n$ and $\gamma_i = 0$ for all $i \geq \log n$. Consequently, since the $\boldsymbol{\Delta}_i$ are mutually independent and $\sum_{i\geq 1} \mathbb{E}[i\boldsymbol{\Delta}_i] = O_n(1)$, (5.43) yields

$$\mathbb{E}\left[\sum_{i\geq 1}i(\boldsymbol{\Delta}_i+\boldsymbol{\gamma}_i)\mathbf{1}\mathcal{Q}_3\right]\leq o_n(1)+4\mathbb{E}\left[\sum_{i\geq 1}i\boldsymbol{\Delta}_i\mathbf{1}\left\{\sum_{i\geq 1}i\boldsymbol{\Delta}_i\geq \ln^3n\right\}\right]=o_n(1). \tag{5.44}$$

Finally, the assertion follows from (5.40), (5.42) and (5.44).

Proof of Lemma 5.2. The first assertion concerning A'' and $A_{n,M}$ follows from Claim 5.17. Concerning A''' and A_{n+1,M^+} , Claim 5.18 shows that it suffices to couple $\operatorname{nul}(A''')\mathbf{1}\mathscr{E}^+$ and $\operatorname{nul}(A_{n+1,M^+})\mathbf{1}\mathscr{E}^+$, because both random variables are bounded by n+1. Indeed, thanks to Claim 5.21 we merely need to couple $\operatorname{nul}(A''')\mathbf{1}\mathscr{E}^+\setminus (\mathscr{U}\cap \mathscr{W})$ and $\operatorname{nul}(A_{n+1,M^+})\mathbf{1}\mathscr{E}^+\setminus (\mathscr{U}\cap \mathscr{W})$, and Claim (5.22) supplies a coupling such that

$$\left|\operatorname{nul}(A''')\mathbf{1}\mathscr{E}^{+} - \operatorname{nul}(A_{n+1,M^{+}})\right|\mathbf{1}\mathscr{E}^{+} \setminus (\mathscr{U} \cap \mathscr{W}) \leq 2\sum_{i>1} i(\Delta_{i} + \gamma_{i})\mathbf{1}\mathscr{E}^{+} \setminus (\mathscr{U} \cap \mathscr{W}). \tag{5.45}$$

Hence, it suffices to show that

$$\mathbb{E}\left[\sum_{i\geq 1}i(\boldsymbol{\Delta}_i+\boldsymbol{\gamma}_i)\mathbf{1}\mathscr{E}^+\setminus(\mathscr{U}\cap\mathscr{W})\right]=o_n(1). \tag{5.46}$$

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

$$\mathbb{E}\left[\sum_{i\geq 1}i(\boldsymbol{\Delta}_i+\boldsymbol{\gamma}_i)\mathbf{1}\mathcal{E}^+\cap\mathcal{W}\setminus\mathcal{U}\right]\leq \left(1-\mathbb{P}\left[\mathcal{U}\mid\mathcal{E}^+\cap\mathcal{W}\right]\right)\ln^4n=o_n(1). \tag{5.47}$$

Finally, the assertion follows from Claim 5.23 and (5.45)–(5.47).

6. Proof of Theorem 1.2

We describe how to extend the proof of [70] to ${\bf G}$. Using the terminology in [70], variable nodes in ${\bf G}$ 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. 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 $\mathbb{E} {\bf d}^{2+\sigma}<\infty$ and $\mathbb{E} {\bf k}^{2+\sigma}<\infty$ by our assumptions on ${\bf d}$ and ${\bf k}$. We claim that

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

Let H_t be the subgraph of G_n obtained after t iterations of the parallel stripping process. Consider Doob's martingale $(\mathbb{E}(H_t \mid e_1, \dots, e_j))_{0 \leq j \leq m}$ where random hyperedges are added in order e_1, \dots, e_m using the configuration model. 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 t-th iteration or not) within its t-neighbourhood. Standard concentration arguments of Azuma's inequality [83, Theorem 2.19] (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 k_1, \dots, k_{d_v} where the k_i s are i.i.d. copies of \hat{k} where

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

Then the *i*-th hyperedge is incident to other $k_i - 1$ vertices (other than v) whose degrees are i.i.d. copies of \hat{a} , where

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

This builds the first neighbourhood of v in T. Iteratively we can build the r-neighbourhood of v in T for any fixed r. It follows from the following claim that the r-neighbourhood of v in G converges in distribution to the r-neighbourhood of T, as $n \to \infty$, for any fixed $r \ge 1$. 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 $x_0x_1...x_\ell = x_0$ 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 $R \ge 1$, $\bigcup_{i \le R} \partial^i(v)$ induces no cycles.

If v survives t iterations of the stripping process then at least 2 hyperedges incident with v survives after t iterations of the stripping process. Let u be a neighbour of v (if there is any) and let x denote the hyperedge containing both u and v. Let ρ_t denote the probability that u is incident with at least 1 hyperedge other than x which survives after t iterations of the stripping process. Note that the degree of u follows the distribution from (6.2). Then, ignoring an o(1) error accounting for the probability of the complement of the evens in Claims 6.1 and 6.2:

and

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

noting that

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

Consequently,

$$\begin{split} \lambda_t &= \sum_{j \geq 2} \mathbb{P}(\boldsymbol{d} = j) \sum_{h \geq 2} \binom{j}{h} \Biggl(\sum_{k_1} \mathbb{P}(\hat{\boldsymbol{k}}_i = k_1) \rho_t^{k_1 - 1} \Biggr)^h \Biggl(\sum_{k_1} \mathbb{P}(\hat{\boldsymbol{k}}_i = k_1) (1 - \rho_t)^{k_1 - 1} \Biggr)^{j - h} \\ &= \sum_{j \geq 2} \mathbb{P}(\boldsymbol{d} = j) \sum_{h \geq 2} \binom{j}{h} \Biggl(\mathbb{E}\rho_t^{\hat{\boldsymbol{k}} - 1} \Biggr)^h \Biggl(1 - \mathbb{E}\rho_t^{\hat{\boldsymbol{k}} - 1} \Biggr)^{j - h} \\ &= \sum_{j \geq 2} \mathbb{P}(\mathcal{D} = j) \Biggl(1 - \left(\frac{K'(\rho_t)}{k} \right)^j - j \frac{K'(\rho_t)}{k} \Biggl(1 - \frac{K'(\rho_t)}{k} \right)^{j - 1} \Biggr) \\ &= 1 - D \Biggl(1 - \frac{K'(\rho_t)}{k} \Biggr) - \frac{K'(\rho_t)}{k} D' \Biggl(\frac{K'(\rho_t)}{k} \Biggr). \end{split}$$

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 non-negative over [0,1]. We also have $\phi(x) = g(x) - x$, where ϕ is given in (7.1). Since $\phi(1) = -D'(0)/d \le 0$, $\phi'(\rho) < 0$ by the hypothesis, and g(x) is nondecreasing in [0,1], it follows that ρ is an attractive fix point of x = g(x). As $\rho_0 = 1$. It follows that $\rho \to \rho$ 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

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

If $\rho=0$ then $\lambda=0$. The case $\rho=0$ of our theorem for \boldsymbol{n}^* follows by letting $I\to\infty$. Since all isolated hyperedges remain in the 2-core, we have $\boldsymbol{m}^*/n=\mathbb{P}(\boldsymbol{k}=0)\boldsymbol{m}/n+o(1)=\frac{d}{L}K(0)+o_{\hat{\epsilon},n}(1)$.

Suppose $\rho > 0$. It is sufficient to show that the 2-core is obtained after removing further $O_n(\hat{\varepsilon}n)$ vertices, following the same approach as [70, Lemma 4]. We briefly sketch it. Following the same argument as before, the probability that a random vertex has degree $j \ge 2$ after I iterations of the stripping process is

$$\sum_{i \geq 2} \mathbb{P}(\boldsymbol{d} = i) \binom{i}{j} (\mathbb{E} \rho_I^{\hat{\boldsymbol{k}} - 1})^j (1 - \mathbb{E} \rho_I^{\hat{\boldsymbol{k}} - 1})^{i - j} = \sum_{i \geq 2} \mathbb{P}(\boldsymbol{d} = i) \binom{i}{j} \left(\frac{K'(\rho_I)}{k} \right)^j \left(1 - \frac{K'(\rho_I)}{k} \right)^{i - j}.$$

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

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

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 $j \ge 2$ vertices is $\gamma_j n + O_n(\hat{\epsilon}n)$, 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 $O_n(\hat{\varepsilon}n)$, and the proportion of remaining hyperedges of size j is

$$\frac{\mathbb{P}(\boldsymbol{k}=j)\rho^{j}}{\sum_{i\geq 1}\mathbb{P}(\boldsymbol{k}=i)\rho^{i}}+O_{n}(\hat{\varepsilon})=\frac{\mathbb{P}(\boldsymbol{k}=j)\rho^{j}}{K(\rho)}+O_{n}(\hat{\varepsilon}),$$

as the probability that a hyperedge survives is proportional to the probability that all of the vertices it contains survive. 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,

$$\begin{split} &\mathbb{E}(X_{t+1} - X_t \mid G_t) \\ &= -1 + \sum_{j \ge 1} \frac{j \mathbb{P}(\boldsymbol{k} = j) \rho^j / K(\rho)}{\rho K'(\rho) / K(\rho)} \cdot (j-1) \cdot \frac{2\gamma_2}{\sum_{i \ge 2} i \gamma_i} + O_n(\hat{\varepsilon}) \\ &= -1 + \frac{1}{\rho K'(\rho)} \left(\sum_{j \ge 1} j (j-1) \mathbb{P}(\boldsymbol{k} = j) \rho^j \right) \frac{2 \cdot \frac{1}{2} (K'(\rho) / k)^2 D''(1 - K'(\rho) / k)}{K'(\rho) d\rho / k} + O_n(\hat{\varepsilon}) \\ &= -1 + \frac{D''(1 - K'(\rho) / k) K''(\rho)}{k d} + O_n(\hat{\varepsilon}). \end{split}$$

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 $j\mathbb{P}(k=j)\rho^j/\rho K'(\rho)$ 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 $2\gamma_2/\sum_{j\geq 2}j\gamma_j$, up to an $O_n(\hat{\varepsilon})$ error. By the assumption that $f'(\rho) < 0$ we have

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

Hence, $\mathbb{E}(X_{t+1}-X_t\mid G_t)<-\delta$ for some $\delta>0$, by making $\hat{\varepsilon}$ sufficiently small (i.e. by choosing sufficiently large I). Then standard Azuma inequality [40, 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 [70, Lemma 4]). The case $\rho>0$ of the theorem follows by

$$\lim_{n\to\infty}\frac{\boldsymbol{m}^*}{n}=\lim_{n\to\infty}\frac{m}{n}\cdot\sum_{i>0}\mathbb{P}(\boldsymbol{k}=j)\rho^j=\frac{d}{k}K(\rho).\quad\Box$$

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 \mathbf{G} , each M is stochastically dominated by $\sum_{j=1}^{N_i(u)} (1+o_n(1))\hat{\mathbf{d}}_j$ where $\hat{\mathbf{d}}_j$ are i.i.d. copies of $\hat{\mathbf{d}}$ whose distribution is given in (6.2). The $o_n(1)$ error is caused by the exposure of $\bigcup_{j \leq i} \partial^j(u)$ which contains $o_n(n)$ vertices by induction. Since $\mathbb{E}\mathbf{d}^{2+\sigma} = O_n(1)$, we have $\hat{d} := \mathbb{E}\hat{\mathbf{d}} = O_n(1)$. 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

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

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 $\hat{\sigma}^i(u)$. Similarly, $\sum_{j=1}^M k_i$ is stochastically dominated by $(1+o_n(1))\sum_{j=1}^M \hat{k}_j$ where \hat{k}_j are i.i.d. copies of \hat{k} whose distribution is defined in (6.1). Let $\hat{k} = \mathbb{E}\hat{k}$. 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 $d_v \le L$. Let k_1, \dots, k_{d_v} be the sizes of the hyperedges incident to ν . Similarly to the proof of Claim 6.1, k_i s are approximated by i.i.d. copies of \hat{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 $1 - \varepsilon$, $\sum_{i=1}^{d_v} k_i \le L$. Inductively, we can make L sufficiently large so that $|\partial^i(v)| \le L$ for all $i \le R$. Let \mathcal{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 \mathcal{E}_i ; (b) two hyperedges in \mathcal{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 $(d_i)_{i \in \partial^i(v)}$ denote the degrees of the vertices in $\partial^i(v)$. The expected number of occurrences of pairs of vertices in

 $\mathbb{E}\left(\sum_{\substack{i:k\in\partial(n)\\2}} \binom{d_j}{2} \binom{d_k}{2} \sum_{h\in[m]} \binom{k_h}{2} O_n(n^{-2})\right) = O_n(n^{-1}) \mathbb{E}\left(\sum_{\substack{i:k\in\partial(n)\\2}} d_j^2 d_k^2\right).$ (6.4)

Note that $|\partial^j(v)| \le L$ 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 1.3

Recall that

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

It is sufficient to prove that if conditions (i) and (ii) are 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: $Var(\mathbf{k}) = 0$. In this case, $\mathbf{k} = k$ 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 (a,b) immediately.

Next consider the case that k=2. Then, $\phi''(\alpha)=-\frac{1}{d}D'''(1-\alpha)<0$ on (0,1) unless $d\leq 2$. Consider the case that $\operatorname{supp} d \cap \mathbb{N}_{\geq 3} \neq \emptyset$. 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 $\mathbf{d} \leq 2$. Then $\phi''(\alpha) = 0$ on [0,1]. Hence $\phi'(\alpha) = \phi'(1) = -1 + \mathbb{P}(d=2)/d < 0$ 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

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

Hence,

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

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

$$\phi'(0) = -1 \qquad \qquad \phi'(1) = -1 + \frac{k-1}{d}D''(0).$$
(7.3)

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 $\Phi'(1) \leq 0$ and thus the supremum of $\Phi(\alpha)$ can only occur at 0 or a stable point. In all of the following sub-cases, 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 $\phi(1)=0$, 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: $Var(\mathbf{k}) = 0$ and $Var(\mathbf{d}) = 0$. In this case $\mathbf{d} = d$. Then $D(\alpha) = \alpha^d$. If $d \ge 3$ then

$$\phi''(\alpha) = \frac{k-1}{d} \alpha^{k-3} \Big((k-2)d(d-1)t^{d-2} - (k-1)d(d-1)(d-2)t^{d-3}(1-t) \Big)$$

$$= (k-1)(d-1)t^{d-3} \alpha^{k-3} \Big((k-2)t - (k-1)(d-2)(1-t) \Big) \quad \text{where } t = 1 - \alpha^{k-1}.$$

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 $d\leq 2$, $\rho=0$ and is the unique maxima. Claims (a) and (b) hold trivially.

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

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

$$h_r(x) = e^x$$
 for all negative integers r. (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,

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

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

$$\frac{k-1}{k-2}(1-t) = \frac{h_{r-2}(\lambda t)}{h_{r-3}(\lambda t)} = 1 - \frac{h_{r-3}(\lambda t) - h_{r-2}(\lambda t)}{h_{r-3}(\lambda t)}.$$
(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: $\mathbf{k} \sim \mathbf{Po}_{>s}(\gamma)$. We must have γ satisfy

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

so that $\mathbb{E} \mathbf{k} = k$. 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,

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

Hence,

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

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

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: $\mathbf{k} \sim \mathbf{Po}_{\geq s}(\gamma)$ and $\operatorname{Var}(\mathbf{d}) = 0$. In this case $\mathbf{d} = d$. Then $D(\alpha) = \alpha^d$. If $d \geq 3$ 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: $\mathbf{k} \sim \mathbf{Po}_{\geq s}(\gamma)$ and $\mathbf{d} \sim \mathbf{Po}_{\geq r}(\lambda)$. 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 [58] to our attention.

REFERENCES

- [1] D. Achlioptas, M. Molloy: The solution space geometry of random linear equations. Random Structures and Algorithms 46 (2015) 197–231.
- [2] M. Aizenman, R. Sims, S. Starr: An extended variational principle for the SK spin-glass model. Phys. Rev. B 68 (2003) 214403.
- [3] R. Alamino, D. Saad: Typical kernel size and number of sparse random matrices over Galois fields: a statistical physics approach. Physical Review E 77 (2008) 061123.
- [4] D. Aldous, M. Steele: The objective method: probabilistic combinatorial optimization and local weak convergence (2003). In: H. Kesten (ed.): Probability on discrete structures, Springer 2004.
- [5] N. Alon, J. Spencer: The probabilistic method. 2nd edition. Wiley (2000).
- [6] P. Ayre, A. Coja-Oghlan, P. Gao, N. Müller: The satisfiability threshold for random linear equations. Combinatorica, in press.
- [7] G. Balakin: On random matrices. Theory Probab. Appl. 12 (1967) 346–353.
- $[8] \ G. \ Balakin: The \ distribution \ of \ random \ matrices \ over \ a \ finite \ field. \ Theory \ Probab. \ Appl. \ 13 \ (1968) \ 631-641.$
- [9] V. Bapst, A. Coja-Oghlan: Harnessing the Bethe free energy. Random Structures and Algorithms 49 (2016) 694–741.
- [10] M. Bayati, D. Gamarnik, P. Tetali: Combinatorial approach to the interpolation method and scaling limits in sparse random graphs. Annals of Probability 41 (2013) 4080–4115.
- [11] J. Blömer, R. Karp, E. Welzl: The rank of sparse random matrices over finite fields. Random Structures and Algorithms 10 (1997) 407-419.
- [12] B. Bollobás: Random graphs. Cambridge University Press (2001).
- [13] B. Bollobás: The evolution of sparse graphs. Graph theory and combinatorics (Cambridge) (1983) 35–57.
- [14] C. Bordenave, M. Lelarge, J. Salez: The rank of diluted random graphs Ann. Probab. 39 (2011) 1097-1121.

- [15] C. Bordenave, M. Lelarge, J. Salez: Matchings on infinite graphs. Probability Theory and Related Fields 157 (2013) 183–208.
- [16] C. Bordenave: A new proof of Friedman's second eigenvalue Theorem and its extension to random lifts. Annales scientifiques de l'école normale supérieure, in press.
- [17] J. Bourgain, V. Vu, P. Wood: On the singularity probability of discrete random matrices. Journal of Functional Analysis 258 (2010) 559-603.
- [18] D. Burshtein, M. Krivelevich, S. Litsyn, G. Miller: Upper bounds on the rate of LDPC codes. IEEE Transactions on Information Theory 48 (2002), 2437–2449.
- [19] J. Cain, N. Wormald: Encores on cores. Electronic journal of combinatorics, 13(1), 81 (2006).
- [20] V. Chvátal: Almost all graphs with 1.44n edges are 3-colorable. Random Structures Algorithms 2(1) (1991) 11–28.
- [21] S. Cocco, O. Dubois, J. Mandler, R. Monasson: Rigorous decimation-based construction of ground pure states for spin glass models on random lattices. Phys. Rev. Lett. **90** (2003) 047205.
- [22] A. Coja-Oghlan, F. Krzakala, W. Perkins and L. Zdeborova: Information-theoretic thresholds from the cavity method. Advances in Mathematics 333 (2018) 694–795.
- [23] A. Coja-Oghlan, W. Perkins: Spin systems on Bethe lattices. arXiv:1808.03440 (2018).
- [24] C. Cooper: The cores of random hypergraphs with a given degree sequence. Random Structures and Algorithms 25(4) (2004) 353–375.
- [25] C. Cooper: On the rank of random matrices. Random Structures and Algorithms 16 (2000) 209-232.
- [26] C. Cooper, A. Frieze, W. Pegden: On the rank of a random binary matrix. arXiv 1806.04988 (2018).
- [27] K. Costello, V. Vu: The rank of random graphs. Random Structures and Algorithms 33 (2008) 269–285.
- [28] K. Costello, V. Vu: On the rank of random sparse matrices. Combinatorics, Probability and Computing 19 (2010) 321-342.
- [29] A. Dembo, E. Lubetzky: Empirical spectral distributions of sparse random graphs. arXiv:1610.05186 (2016).
- [30] M. Dietzfelbinger, A. Goerdt, M. Mitzenmacher, A. Montanari, R. Pagh, M. Rink: Tight thresholds for cuckoo hashing via XORSAT. Proc. 37th ICALP (2010) 213–225.
- [31] O. Dubois, J. Mandler: The 3-XORSAT threshold. Proc. 43rd FOCS (2002) 769–778.
- [32] R. Durrett: Probability theory and examples, 3rd edition (2005).
- [33] P. Erdös, A. Rényi: On random matrices. Magyar Tud. Akad. Mat. Kutató Int. Közl. 8 (1963) 455-461.
- [34] D. Fernholz, V. Ramachandran: The giant k-core of a random graph with a specified degree sequence. Manuscript, 2003.
- [35] D. Fernholz, V. Ramachandran: Cores and connectivity in sparse random graphs. The University of Texas at Austin, Department of Computer Sciences, technical report TR-04-13 (2004).
- [36] S. Franz, M. Leone: Replica bounds for optimization problems and diluted spin systems. J. Stat. Phys. 111 (2003) 535-564.
- [37] J. Friedman: A proof of Alon's second eigenvalue conjecture and related problems. Memoirs of the AMS (2008).
- [38] J. Fulman, L. Goldstein: SteinâĂŹs method and the rank distribution of random matrices over finite fields. Annals of Probability **43** (2015) 1274–1314.
- [39] R. Gallager: Low-density parity check codes. IRE Trans. Inform. Theory 8 (1962) 21–28.
- [40] P. Gao, M. Molloy: The stripping process can be slow: part I. arXiv:1501.02695 (2015).
- [41] A. Giurgiu, N. Macris, R. Urbanke: Spatial coupling as a proof technique and three applications. IEEE Transactions on Information Theory 62 (2016) 5281–5295.
- [42] A. Goerdt, L. Falke: Satisfiability thresholds beyond *k*-XORSAT. Proc. 7th International Computer Science Symposium in Russia (2012) 148–159.
- [43] F. Guerra: Broken replica symmetry bounds in the mean field spin glass model. Communications in Mathematical Physics 233 (2003) 1–12.
- [44] M. Ibrahimi, Y. Kanoria, M. Kraning, A. Montanari: The set of solutions of random XORSAT formulae. Annals of Applied Probability 25 (2015) 2743–2808.
- [45] S. Janson, M. Luczak: A simple solution to the k-core problem. Random Structures Algorithms 30(1-2) (2007) 50-62.
- [46] S. Janson, T. Łuczak, A. Ruciński: Random graphs. Wiley (2000).
- [47] Y. Kabashima, D. Saad: Statistical mechanics of error correcting codes. Europhys. Lett. 45 (1999) 97-103.
- [48] J. Kahn; J. Komloś, E. Szemerédi: On the probability that a random ±1-matrix is singular. Journal of the AMS 8 (1995) 223–240.
- [49] J. Kim: Poisson cloning model for random graphs. International Congress of Mathematicians. Vol. III, Eur. Math. Soc., Zürich (2006) 873–897.
- [50] V. Kolchin: Random graphs and systems of linear equations in finite fields. Random Structures and Algorithms 5 (1995) 425–436.
- [51] V. Kolchin, V. Khokhlov: On the number of cycles in a random non-equiprobable graph. Discrete Math. Appl. 2 (1992) 109–118.
- [52] V. Kolchin: Consistency of a system of random congruences. Discrete Math. Appl. 3 (1993) 103-113.
- [53] J. Komlós: On the determinant of (0,1) matrices. Studia Sci. Math. Hungar. 2 (1967) 7–21.
- [54] 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.
- [55] I. Kovalenko, A. Levitskaya, M. Savchuk: Selected problems of probabilistic combinatorics. Naukova Dumka, Kiev (1986).
- [56] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, P. Zhang: Spectral redemption in clustering sparse networks. Proc. National Academy of Sciences 110 (2013) 20935–20940.
- [57] S. Kumar, A. Young, N. Macris, H. Pfister: Threshold saturation for spatially-coupled LDPC and LDGM Codes on BMS channels. IEEE Transactions on Information Theory 60 (2014) 7389–7415
- [58] M. Lelarge: Bypassing correlation decay for matchings with an application to XORSAT. Proc. IEEE Information Theory Workshop (2013)
- [59] M. Lelarge, M. Oulamara: Replica bounds by combinatorial interpolation for diluted spin systems. Journal of Statistical Physics (2018)
- [60] A. Levitskaya: Theorems on invariance for the systems of random linear equations over an arbitrary finite ring. Soviet Math. Dokl. 263 (1982) 289–291.

- [61] A. Levitskaya: The probability of consistency of a system of random linear equations over a finite ring. Theory Probab. Appl. **30** (1985) 339–350.
- [62] T. Luczak: Size and connectivity of the k-core of a random graph. Discrete Math., 91(1) (1991) 61-68.
- [63] T. Luczak: Sparse random graphs with a given degree sequence. Random graphs, Vol. 2 (Poznań, 1989), Wiley-Intersci. Publ. Wiley, New York, (1992) 165–182.
- [64] K. Luh, S. Meehan, H. Nguyen: Random matrices over finite fields: methods and results. arXiv:1907.02575 (2019).
- [65] C. Méasson, A. Montanari, R. Urbanke: Maxwell's constructions: the hidden bridge between maximum-likelihood and iterative decoding. IEEE Transactions on Information Theory 54 (2008) 5277–5307.
- [66] M. Mehta: Random matrices. Academic press (2004).
- [67] M. Mézard, A. Montanari: Information, physics and computation. Oxford University Press 2009.
- [68] M. Mézard, F. Ricci-Tersenghi, R. Zecchina: Two solutions to diluted *p*-spin models and XORSAT problems. Journal of Statistical Physics 111 (2003) 505–533.
- [69] G. Miller, G. Cohen: The rate of regular LDPC codes. IEEE Transactions on Information Theory 49 (2003) 2989–2992.
- [70] M. Molloy: Cores in random hypergraphs and Boolean formulas. Random Structures and Algorithms 27 (2005) 124–135.
- [71] A. Montanari: Estimating random variables from random sparse observations. European Transactions on Telecommunications 19(4) (2008) 385–403.
- [72] A. Montanari: Tight bounds for LDPC and LDGM codes under MAP decoding. IEEE Transactions on Information Theory 51 (2005) 3221-3246.
- [73] D. Panchenko: Spin glass models from the point of view of spin distributions. Annals of Probability 41 (2013) 1315–1361.
- [74] B. Pittel, J. Spencer, N. Wormald: Sudden emergence of a giant k-core in a random graph. J. Combin. Theory Ser. B, 67(1) (1996) 111–151.
- [75] B. Pittel, G. Sorkin: The satisfiability threshold for k-XORSAT. Combinatorics, Probability and Computing 25 (2016) 236–268.
- [76] P. Raghavendra, N. Tan: Approximating CSPs with global cardinality constraints using SDP hierarchies. Proc. 23rd SODA (2012) 373–387.
- [77] T. Richardson, R. Urbanke: Modern coding theory. Cambridge University Press (2008).
- [78] O. Riordan: The k-core and branching processes. Combin. Probab. Comput.17(1) (2008) 111-136.
- [79] T. Tao, V. Vu: On the singularity probability of random Bernoulli matrices. Journal of the AMS 20 (2007) 603–628.
- [80] K. Tikhomirov: Singularity of random Bernoulli matrices. arXiv:1812.09016.
- [81] V. Vu: Combinatorial problems in random matrix theory. Proc. International Congress of Mathematicians IV (2014) 489-508.
- [82] E. Wigner: Characteristic vectors of bordered matrices with infinite dimensions. Annals of Mathematics 62 (1955) 548-564.
- [83] N. Wormald: Models of random regular graphs. London Mathematical Society Lecture Note Series (1999) 239-298.
- [84] L. Zdeborová, F. Krzakala: Statistical physics of inference: thresholds and algorithms. Advances in Physics 65 (2016) 453-552.

APPENDIX A. PROOF OF LEMMA 1.8

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

$$\mathbb{P}\left[\mathcal{M}\right] = 1 - o_n(1/n). \tag{A.1}$$

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 $\mathbb{P}\left[\forall j \leq L : |Q_j - s\mathbb{P}\left[\lambda = j\right]| > \sqrt{n} \ln n\right] = o_n(1/n)$. Hence, by Bayes' rule,

$$\mathbb{P}\left[\exists j \le L : |Q_j - s\mathbb{P}\left[\lambda = j\right]| > \sqrt{n} \ln n \,|\, \mathcal{M}\right] = o_n(1/n). \tag{A.2}$$

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

$$R_h = \sum_{j \geq 1} Q_j \mathbf{1} \{ L(1+\eta)^{h-1} < j \leq L(1+\eta)^h \wedge n / \ln^9 n \}, \quad \bar{R}_h = s \sum_{j \geq 1} \mathbb{P} \left[\boldsymbol{\lambda} = j \right] \mathbf{1} \{ L(1+\eta)^{h-1} < j \leq L(1+\eta)^h \wedge n / \ln^9 n \}.$$

Then the Chernoff bound and Bayes' rule yield

$$\mathbb{P}\left[\exists h \in \mathcal{H} : \left| R_h - \bar{R}_h \right| > \eta \bar{R}_h + \ln^2 n \mid \mathcal{M} \right] = o_n(1/n). \tag{A.3}$$

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

$$\begin{split} \frac{1}{s} \sum_{i=1}^{s} \lambda_{i} &\leq \sum_{j=1}^{s} j Q_{j} / s + \sum_{h \in \mathcal{H}} (1 + \eta)^{h} L R_{h} / s \\ &= o_{n}(1) + \mathbb{E} \left[\lambda \mathbf{1} \{ \lambda \leq L \} \right] + \sum_{h \in \mathcal{H}} (1 + \eta)^{h+1} (\bar{R}_{h} + (\ln^{2} n)) / s \leq \mathbb{E} \left[\lambda \mathbf{1} \right] + \delta / 2 + o_{n}(1). \end{split}$$

Similarly, $\frac{1}{s} \sum_{i=1}^{s} \lambda_i \ge \mathbb{E}[\lambda 1] - \delta/2 + o_n(1)$. Thus, the assertion follows from (A.1)–(A.3)

APPENDIX B. STOCHASTIC VS. LINEAR INDEPENDENCE

A precursor of Proposition 2.3 for finite field was obtained in [6, 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 $\sigma = \sigma_A \in \mathbb{F}^n$ denote a sample from μ_A . The stochastic independence statement reads as follows.

Lemma B.1 ([6, Lemma 3.1]). For any $\delta > 0$, $\ell > 0$ and for any finite field \mathbb{F} there exists $\mathcal{T} = \mathcal{T}(\delta, \ell, \mathbb{F}) > 0$ such that for any matrix A over \mathbb{F} the following is true. Choose $\theta \in [\mathcal{T}]$ uniformly at random. Then with probability at least $1 - \delta$ the matrix $A[\theta]$ satisfies

$$\sum_{I\subseteq[n]:|I|=\ell} \max_{\tau\in\mathbb{F}^I} \left| \mu_{A[\boldsymbol{\theta}]} \left(\{ \forall i \in I : \boldsymbol{\sigma}_i = \boldsymbol{\tau}_i \} \right) - \prod_{i\in I} \mu_{A[\boldsymbol{\theta}]} \left(\{ \boldsymbol{\sigma}_i = \boldsymbol{\tau}_i \} \right) \right| < \delta n^{\ell}. \tag{B.1}$$

In words, for most sets I of ℓ coordinates the joint distribution of the coordinates $(\sigma_i)_{i \in I}$ is close to a product distribution in total variation distance. Furthermore, the number θ of rows that we add to A is bounded in terms of ε , ℓ only; i.e., θ 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 [22].

The following lemma shows that how Proposition 2.3 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

$$\mu_A\left(\{\forall i \in I : \boldsymbol{\sigma}_i = \boldsymbol{\tau}_i\}\right) = \prod_{i \in I} \mu_A\left(\{\boldsymbol{\sigma}_i = \boldsymbol{\tau}_i\}\right) \qquad \qquad \text{for all } \boldsymbol{\tau} \in \mathbb{F}^I. \tag{B.2}$$

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

$$\begin{split} \mu_A\left(\{\forall i \in I: \pmb{\sigma}_i = \tau_i\}\right) &= \mathbf{1}\left\{\forall i \in I \cap \mathfrak{F}(A): \tau_i = 0\right\} \mu_A\left(\{\forall i \in I \setminus \mathfrak{F}(A): \pmb{\sigma}_i = \tau_i\}\right), \\ \prod_{i \in I} \mu_A\left(\{\pmb{\sigma}_i = \tau_i\}\right) &= \mathbf{1}\left\{\forall i \in I \cap \mathfrak{F}(A): \tau_i = 0\right\} \prod_{i \in I \setminus \mathfrak{F}(A)} \mu_A\left(\{\pmb{\sigma}_i = \tau_i\}\right), \end{split}$$

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 (B.2) under the assumption that $I \subseteq [n] \setminus \mathfrak{F}(A)$ is no relation of A.

To prove this statement let $N=\operatorname{nul}(A)$ 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 ξ_1,\ldots,ξ_N and let Ξ_1,\ldots,Ξ_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 (B.2) it suffices to prove that the projection of this homomorphism to the I-rows, i.e., the map $z\in\mathbb{F}^N\mapsto(\Xi_iz)_{i\in I}$ is surjective. Equivalently, we need to show that

$$\operatorname{rk}(\Xi_i)_{i \in I} = |I|. \tag{B.3}$$

Assume for contradiction that (B.3) 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 $(yA)_j = 1$ $\{i \in I\}$ z_i for all $j \in [n]$. Hence, $\emptyset \neq \text{supp}(yA) \subseteq I$. 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 2.3 and Lemma B.2. Indeed, the proof of Proposition 2.3 renders the explicit bound $\mathcal{T} = \lceil 4\ell^3/\delta^4 \rceil + 1$ on the number of coordinates that need to get pegged. By comparison, the stochastic approach via the arguments from [6, 9] leads to a value of \mathcal{T} that is exponential in ℓ (although it may be possible to improve this estimate via probabilistic arguments).

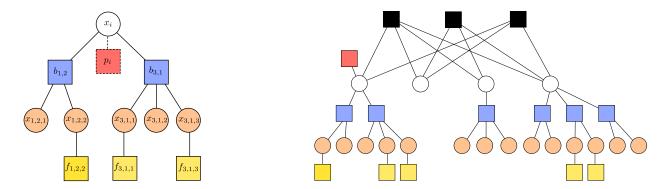


FIGURE 2. Left: sketch of the component of x_i at t = 0; the check p_i is present iff $i \le \theta$. Right: sketch of the factor graph $G_{\varepsilon}(t)$ for $0 < t < m_{\varepsilon,n}$, with the $a_{i,j}$ coloured black and the other colours as in the left figure.

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

The '≤'-inequality in (1.3) was previously proved by Lelarge [58], who derived the bound from the Leibniz determinant formula and the formula for the matching number of random bipartite graphs from [14]. 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, the interpolation method from spin glass theory [36, 43], 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 w.h.p.

$$\operatorname{nul}(A)/n \ge \max_{\alpha \in [0,1]} \Phi(\alpha) + o_n(1). \tag{C.1}$$

C.1. **The interpolation method.** The basic idea behind the interpolation method is to construct a family of random matrices $A_{\varepsilon}(t)$ parametrised by 'time' t. At $t = m_{\varepsilon,n}$ we obtain precisely the matrix $A_{\varepsilon,n}$. At the other extreme, $A_{\varepsilon}(0)$ 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., [10, 22, 36, 73]), 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 $\mathcal{T} = \mathcal{T}(\varepsilon) \ge 0$ and a real $\beta \in [0,1]$, which, in order to obtain the optimal bound, we choose such that

$$\Phi(\beta) = \max_{\alpha \in [0,1]} \Phi(\alpha). \tag{C.2}$$

Further, let $\mathbf{m}_{\varepsilon,n} \sim \operatorname{Po}((1-\varepsilon)dn/k)$. Also let $(\mathbf{k}_i, \mathbf{k}_i', \mathbf{k}_i'')_{i\geq 1}$ and $(\mathbf{d}_i)_{i\geq 1}$ be copies of \mathbf{k} and \mathbf{d} , respectively, mutually independent and independent of $\mathbf{m}_{\varepsilon,n}$. Additionally, choose $\mathbf{\theta} \in [\mathcal{F}]$ 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 $G_{\varepsilon}(t)$ has variable nodes

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

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

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

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

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

For each matching edge $\{(x_i, s), (a_j, t)\} \in \Gamma_{\varepsilon}(t)$ insert an edge between x_i and a_j into the Tanner graph and for each $\{(x_i, s), b_{j,h}\} \in \Gamma_{\varepsilon}(t)$ insert an edge between x_i and $b_{j,h}$. Thus, $G_{\varepsilon}(t)$ may contain multi-edges. Further, add an edge

between x_i and p_i for $i = 1, ..., \theta$ and add an edge between $x_{i,j,h}$ and $b_{i,j}$ for each $h \in [k'_i - 1]$ as well as an edge between every $x_{i,j,h} \in \mathscr{F}_t$ and the check $f_{i,j,h}$. Finally, let $A_{\varepsilon}(t)$ be the random matrix induced by $G_{\varepsilon}(t)$. Formally, with the rows indexed by the check nodes and the columns indexed by the variable nodes, we let

$$(A_{\varepsilon}(t))_{p_{i},x_{j}} = \mathbf{1}\left\{i = j\right\}$$

$$(i \in [\theta], j \in [n]),$$

$$(A_{\varepsilon}(t))_{a_{i},x_{j}} = \chi_{\zeta_{i},\xi_{j}} \sum_{u=1}^{k_{i}} \sum_{v=1}^{d_{j}} \mathbf{1}\left\{\{(x_{j},v),(a_{i},u) \in \Gamma_{\varepsilon}(t)\}\right\}$$

$$(i \in [t], j \in [n]),$$

$$(A_{\varepsilon}(t))_{b_{h,i},x_{j}} = \mathbf{1}\left\{x_{j} \in \partial_{G_{\varepsilon}(t)}b_{h,i}\right\}$$

$$(h \in [m_{\varepsilon,n}-t], j \in [n]),$$

$$(A_{\varepsilon}(t))_{b_{h,i},x_{u,v,w}} = \mathbf{1}\left\{h = u, i = v\right\}$$

$$(h, u \in [m_{\varepsilon,n}-t], i \in [k'_{h}], v \in [k'_{u}-1]),$$

$$(h, u \in [m_{\varepsilon,n}-t], i \in [k'_{h}], j \in [vk'_{h}-1],$$

$$v \in [k'_{u}], w \in [k'_{u}-1]).$$

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

The semantics is as follows. The checks a_i will play exactly the same role as before, i.e., each is adjacent to \mathbf{k}_i of the variable nodes x_1, \ldots, x_n w.h.p. 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 $\mathbf{k}'_i - 1$ variable nodes $x_{i,j,h}$, $h \in [\mathbf{k}'_i - 1]$. 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, i.e., $f_{i,j,h}$ 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_{θ} just freeze x_1, \ldots, x_{θ} .

For t=1 the Tanner graph contains $\mathbf{m}_{\varepsilon,n} \sim \text{Po}((1-\varepsilon)dn/k)$ 'real' checks a_i and none of the checks $b_{i,j}$ or $f_{i,j,h}$. In effect, $A_{\varepsilon}(1)$ is distributed precisely as A_{ε} 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 $i \in [\theta]$ there is a check p_i freezing x_i to zero as well. Thus, $A_{\varepsilon}(0)$ is a block diagonal matrix consisting of n blocks, one for each component. In effect, the rank of $A_{\varepsilon}(0)$ 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 2.

We are going to trace the nullity of $A_{\varepsilon}(t)$ 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 $A_{\varepsilon}(0)$ by comparison to $A_{\varepsilon}(m_{\varepsilon,n})$. Thus, let

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

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

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

The next proposition provides monotonicity.

Proposition C.2. For any $\varepsilon > 0$ there exists $\mathcal{T} = \mathcal{T}(\varepsilon) > 0$ such that with probability $1 - o_n(1/n)$ uniformly for all $0 \le t < m_{\varepsilon,n}$ we have $\mathbb{E}[\mathcal{N}_{t+1} + \mathcal{Y}_{t+1} \mid m_{\varepsilon,n}] \ge \mathbb{E}[\mathcal{N}_t + \mathcal{Y}_t \mid m_{\varepsilon,n}] + o_{\varepsilon,n}(1)$.

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 $\frac{1}{n}\mathbb{E}[\operatorname{nul}(A_{\varepsilon})] \ge \max_{\alpha \in [0,1]} \Phi(\alpha) + o_{\varepsilon,n}(1)$.

Proof. Proposition C.2 implies that

$$\mathbb{E}[\operatorname{nul} \mathbf{A}_{\varepsilon,n}] = \mathbb{E}[\operatorname{nul} \mathbf{A}_{\varepsilon}(\mathbf{m}_{\varepsilon,n})] = \mathbb{E}[\mathcal{N}_{\mathbf{m}_{\varepsilon,n}}] = \mathbb{E}[\mathcal{N}_{\mathbf{m}_{\varepsilon,n}} + \mathcal{Y}_{\mathbf{m}_{\varepsilon,n}}] - \mathbb{E}[\mathcal{Y}_{\mathbf{m}_{\varepsilon,n}}]$$

$$\geq \mathbb{E}[\mathcal{N}_0 + \mathcal{Y}_0] - \mathbb{E}[\mathcal{Y}_{\mathbf{m}_{\varepsilon,n}}] + o_{\varepsilon}(n) = \mathbb{E}[\mathcal{N}_0] - \mathbb{E}[\mathcal{Y}_{\mathbf{m}_{\varepsilon,n}}] + o_{\varepsilon,n}(n). \tag{C.3}$$

Further, by Proposition C.1,

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

Hence, (C.2) yields

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

and the assertion follows from (C.3).

Combining Proposition 2.5, Proposition 2.7 and Corollary C.3 and the standard concentration for nul A_{ε} from Lemma 4.7 completes the proof of (C.1). We proceed to prove Propositions C.1 and C.2.

C.2. **Proof of Proposition C.1.** Each component of $G_{\mathcal{E}}(0)$ contains precisely one of the variable nodes x_1, \ldots, x_n . In effect, $A_{\mathcal{E}}(0)$ 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 \mathbf{B}_s representing the connected component of x_s . Indeed, because $\sum_{s=1}^{n} \left| \partial^2 x_s \right| = \sum_{i \leq \mathbf{m}'_{\mathcal{E}}(0)} \mathbf{k}'_i(\mathbf{k}'_i - 1)$ and $\sum_{s=1}^{n} \left| \partial^2 x_s \cap \mathscr{F}_0 \right| = |\mathscr{F}_0|$ we have

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

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

$$\mathbb{E}[\mathbf{N}_s] = \begin{cases} dK'(\beta)/k + D(1 - K'(\beta)/k) - d + o_{\varepsilon}(1) & \text{if } s > \boldsymbol{\theta}, \\ O_n(1) & \text{otherwise.} \end{cases}$$
(C.4)

In fact, the second case in (C.4) simply follows from $N_s \le d_s$ and $\mathbb{E}[d_s] = O_n(1)$ for all s.

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

$$\mathbb{E}[|\mathbf{N}_s|\mathbf{1}\{\mathbf{d}_s > \varepsilon^{\xi - 1/2}\}] = o_{\varepsilon,n}(1). \tag{C.5}$$

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

$$\mathbb{E}\left[\Xi\right] \le \frac{dn}{k} \mathbb{E}\left[k\mathbf{1}\{k \ge \varepsilon^{-8}\}\right] = nO_{\varepsilon,n}(\varepsilon^{8}),\tag{C.6}$$

while $M'_{j} \sim (1 - \varepsilon) dn \mathbb{P}\left[k = j\right] / k$ for all $j \le \varepsilon^{-8}$ w.h.p. by Chebyshev's inequality. Hence, introducing the event

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

we obtain from (C.5) and (C.6) that

$$\mathbb{E}[N_s] = \mathbb{E}[N_s \mathbf{1}\mathscr{E}_s] + o_{\varepsilon,n}(1). \tag{C.7}$$

With $\gamma \leq d_s$ the actual degree of x_s in $G_{\varepsilon}(s)$, let $\kappa_1, \dots, \kappa_{\gamma}$ be the degrees of the checks adjacent to x_s . We claim that given \mathscr{E}_s and d_s ,

$$d_{\text{TV}}((\boldsymbol{\kappa}_1, \dots, \boldsymbol{\kappa}_{\gamma}), (\hat{\boldsymbol{k}}_1, \dots, \hat{\boldsymbol{k}}_{\boldsymbol{d}_s})) = o_{\varepsilon, n}(\varepsilon^{1/2}). \tag{C.8}$$

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

$$\sum_{j>3} j \mathbf{M}'_j \ge (1-2\varepsilon) dn,$$

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

$$j\boldsymbol{M}_{j}^{\prime}/\sum_{h\geq3}h\boldsymbol{M}_{h}^{\prime}=j\mathbb{P}\left[\boldsymbol{k}=j\right]/k+o_{n}(1)=\mathbb{P}\left[\boldsymbol{\hat{k}}=j\right]+o_{n}(1).$$

These events are asymptotically independent for the different clones. Thus, we obtain (C.8).

Finally, we can easily compute N_s given the vector $(\kappa_1,...,\kappa_{\gamma})$. The matrix B_s has fairly simple structure. The first γ rows have a non-zero entry in the first column representing x_s . Additionally, for $i=1,...,\gamma$ the ith row contains κ_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

$$\mathbb{E}[\operatorname{rk}(\boldsymbol{B}_{s}) \mid \boldsymbol{\kappa}_{1}, \dots, \boldsymbol{\kappa}_{\gamma}] = \sum_{i=1}^{\gamma} (1 - \beta^{\boldsymbol{\kappa}_{i}-1}) + |\mathscr{F}_{0} \cap \partial^{2} x_{s}| + 1 - \prod_{i=1}^{\gamma} (1 - \beta^{\boldsymbol{\kappa}_{i}-1}). \tag{C.9}$$

To see this, let us first compute the rank of the matrix \mathbf{B}_s' without the first column. Then row $i \in [\gamma]$ 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 β^{κ_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 \mathbf{B}_s , the first column adds to the rank unless non of the first γ rows of \mathbf{B}_s' gets zeroed out completely, an event that has probability $\prod_{i=1}^{\gamma} (1-\beta^{\kappa_i-1})$. Since

$$\mathbb{E}[N_{s} \mid \boldsymbol{\kappa}_{1}, \dots, \boldsymbol{\kappa}_{\gamma}] = 1 + \sum_{i=1}^{\gamma} (\boldsymbol{\kappa}_{i} - 1) - \mathbb{E}[\operatorname{rk}(\boldsymbol{B}_{s}) \mid \boldsymbol{\kappa}_{1}, \dots, \boldsymbol{\kappa}_{\gamma}] - \mathbb{E}[\left|\partial^{2} x_{s}\right| - \left|\partial^{2} x_{s} \cap \mathscr{F}_{0}\right| \mid \boldsymbol{\kappa}_{1}, \dots, \boldsymbol{\kappa}_{\gamma}]$$

$$= 1 - \mathbb{E}[\operatorname{rk}(\boldsymbol{B}_{s}) \mid \boldsymbol{\kappa}_{1}, \dots, \boldsymbol{\kappa}_{\gamma}] + \mathbb{E}\left[\left|\partial^{2} x_{s} \cap \mathscr{F}_{0}\right| \mid \boldsymbol{\kappa}_{1}, \dots, \boldsymbol{\kappa}_{\gamma}\right],$$

substituting (C.9) in yields

$$\mathbb{E}[N_{s} \mid \kappa_{1}, \dots, \kappa_{\gamma}] = \prod_{i=1}^{\gamma} (1 - \beta^{\kappa_{i}-1}) - \sum_{i=1}^{\gamma} (1 - \beta^{\kappa_{i}-1}). \tag{C.10}$$

Combining (C.7), (C.8) and (C.10) completes the proof.

C.3. **Proof of Proposition C.2.** To couple the random variables \mathcal{N}_{t+1} and \mathcal{N}_t we need to investigate short linear relations among the cavities, i.e., the clones from $\bigcup_{i=1}^n \{x_i\} \times [\boldsymbol{d}_i]$ that are not incident to an edge of $\Gamma_{\varepsilon}(t)$. 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 $y_1, y_2...$ be independent samples from P_t .

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

$$\mathbb{P}\left[y_1, ..., y_\ell \text{ form a proper relation}\right] < \delta.$$

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

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

$$\mathbb{P}\left[\mathbf{y}_{1}=y_{1},\ldots,\mathbf{y}_{\ell}=y_{\ell}\mid A_{\varepsilon}(t)\right]\leq\mathbb{P}\left[\mathbf{x}_{1}=y_{1},\ldots,\mathbf{x}_{\ell}=y_{\ell}\mid A_{\varepsilon}(t)\right](2L/\varepsilon)^{\ell}+\delta^{2}.$$

Consequently, because the distribution of $G_{\varepsilon}(t) - \{p_1, ..., p_{\theta}\}$ is invariant under permutations of $x_1, ..., x_n$, Remark 3.5 shows that $\mathbb{P}\left[x_1 = y_1, ..., x_{\ell} = y_{\ell} \mid A_{\varepsilon}(t)\right] < \delta(\varepsilon/(2L))^{\ell}/2$, provided that $\mathcal{T} = \mathcal{T}(\delta, \ell)$ is large enough.

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

$$\mathcal{N}_t' = \operatorname{nul} A_{\varepsilon}'(t) + |\mathcal{F}_{t+1}| - \sum_{i=1}^{m_{\varepsilon,n}-t-1} k_i'(k_i'-1).$$

Then clearly

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

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

Lemma C.5. We have
$$\mathbb{E}\left|\mathbb{E}[\mathcal{N}_{t+1} - \mathcal{N}_t' \mid A_{\varepsilon}(t)', m_{\varepsilon,n}] - (K(\alpha) - 1)\right| = o_{\varepsilon,n}(1)$$
.

Proof. The random matrix $A_{\varepsilon}(t+1)$ is obtained from $A'_{\varepsilon}(t)$ by inserting a new random check a_{t+1} . Pick $\zeta = \zeta(\varepsilon) > 0$ small enough and $\delta = \delta(\zeta) > 0$ smaller still. Since $|\operatorname{nul}(A'_{\varepsilon}(t)) - \operatorname{nul}(A_{\varepsilon}(t+1))| \le 1$ and $\mathbb{E}[k^2] = O_{\varepsilon,n}(1)$ we may condition on the event that $k_{t+1} \le \varepsilon^{-1}$. Similarly, Lemma C.4 shows that we may assume that the set \mathscr{X} of variables of $G'_{\varepsilon}(t)$ 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 2.4 yields

$$\mathbb{E}[\mathcal{N}_{t+1} - \mathcal{N}_t \mid A_{\varepsilon}(t)', \boldsymbol{m}_{\varepsilon,n}] = \mathbb{E}[\operatorname{nul}(\boldsymbol{A}_{\varepsilon}(t+1)) - \operatorname{nul}(\boldsymbol{A}'_{\varepsilon}(t)) \mid \boldsymbol{A}'_{\varepsilon}(t), \boldsymbol{m}_{\varepsilon,n}]$$

$$= \mathbb{E}\left[\boldsymbol{\alpha}^{\boldsymbol{k}_{t+1}} - 1 \mid \boldsymbol{A}'_{\varepsilon}(t), \boldsymbol{m}_{\varepsilon,n}\right] + o_{\varepsilon}(1) = K(\boldsymbol{\alpha}) - 1 + o_{\varepsilon,n}(1),$$

as claimed. \Box

Lemma C.6. Let
$$Q(\alpha, \beta) = \mathbb{E}[k(\alpha\beta^{k-1} - 1)]$$
 for $\alpha \in [0, 1]$. Then $\mathbb{E}[\mathbb{E}[\mathcal{N}_t - \mathcal{N}_t' \mid A_{\varepsilon}'(t), m_{\varepsilon, n}] - Q(\alpha, \beta)] = o_{\varepsilon, n}(1)$.

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

$$|\mathcal{N}_t - \mathcal{N}_t'| \le \mathbf{k}'_{\mathbf{m}_{\varepsilon} - t - 1}$$

and $\mathbb{E}[k^2] = O_{\varepsilon,n}(1)$ we may condition on the event that $k'_{m_{\varepsilon}-t-1} \leq \varepsilon^{-1}$. In effect, Lemma C.4 shows that we may assume the set \mathscr{X} of cavities adjacent to the new checks $b_{m_{\varepsilon,n}-t-1,h}$ 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 $\mathrm{Bin}(k'_{m_{\varepsilon,n}-t-1},\alpha)$ in total variation. Therefore, Lemma 2.4 shows that

$$\mathbb{E}\left[\mathcal{N}_{t} - \mathcal{N}_{t}' \mid A_{\varepsilon,n}'(t), \boldsymbol{m}_{\varepsilon,n}\right] = \mathbb{E}\left[\operatorname{nul}(\boldsymbol{A}_{\varepsilon}(t)) - \operatorname{nul}(\boldsymbol{A}_{\varepsilon}'(t)) - \boldsymbol{k}_{\boldsymbol{m}_{\varepsilon,n}-t-1}'(\boldsymbol{k}_{\boldsymbol{m}_{\varepsilon,n}-t-1}' - 1) + |\mathcal{F}'| \mid A_{\varepsilon}'(t), \boldsymbol{m}_{\varepsilon,n}\right] = Q(\boldsymbol{\alpha}, \boldsymbol{\beta}) + o_{\varepsilon,n}(1),$$

as claimed. \Box

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

Proof. This is the result of a straightforward calculation.

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

$$\mathbb{E}[\mathcal{N}_{t+1} + \mathcal{Y}_{t+1}] - \mathbb{E}[\mathcal{N}_t + \mathcal{Y}_t] = \mathbb{E}\left[\boldsymbol{\alpha}^{k} - 1 - \boldsymbol{k}(\boldsymbol{\alpha}\boldsymbol{\beta}^{k-1} - 1) + (\boldsymbol{k} - 1)(\boldsymbol{\beta}^{k} - 1)\right] + o_{\varepsilon,n}(1). \tag{C.11}$$

Since $x^k - kxy^{k-1} + (k-1)y^k \ge 0$ for all $k \ge 1$, $x, y \in [0, 1]$, the assertion follows from (C.11).

AMIN COJA-OGHLAN, acoghlan@math.uni-frankfurt.de, GOETHE UNIVERSITY, MATHEMATICS INSTITUTE, 10 ROBERT MAYER ST, FRANK-FURT 60325, GERMANY.

ALPEREN A. ERGÜR, aergur@cs.cmu.edu, CARNEGIE MELLON UNIVERSITY, SCHOOL OF COMPUTER SCIENCE, 7225 GATES HILLMAN CENTER, PITTSBURGH, PA, 15213, USA.

PU GAO, p3gao@uwaterloo.ca, DEPARTMENT OF COMBINATORICS AND OPTIMIZATION UNIVERSITY OF WATERLOO, CANADA.

SAMUEL HETTERICH, hetterich@math.uni-frankfurt.de, GOETHE UNIVERSITY, MATHEMATICS INSTITUTE, 10 ROBERT MAYER ST, FRANK-FURT 60325, GERMANY.

MAURICE ROLVIEN, rolvien@math.uni-frankfurt.de, GOETHE UNIVERSITY, MATHEMATICS INSTITUTE, 10 ROBERT MAYER ST, FRANK-FURT 60325, GERMANY.