THE SET OF SOLUTIONS OF RANDOM XORSAT FORMULAE[1]

By Morteza Ibrahimi, Yash Kanoria[2],
Matt Kraning and Andrea Montanari

Urban Engines, Columbia Business School, Qadium, Inc.
and Stanford University

The XOR-satisfiability (XORSAT) problem requires finding an assignment of n Boolean variables that satisfy m exclusive OR (XOR) clauses, whereby each clause constrains a subset of the variables. We consider random XORSAT instances, drawn uniformly at random from the ensemble of formulae containing n variables and m clauses of size k. This model presents several structural similarities to other ensembles of constraint satisfaction problems, such as k-satisfiability (k-SAT), hypergraph bicoloring and graph coloring. For many of these ensembles, as the number of constraints per variable grows, the set of solutions shatters into an exponential number of well-separated components. This phenomenon appears to be related to the difficulty of solving random instances of such problems.

We prove a complete characterization of this clustering phase transition for random k-XORSAT. In particular, we prove that the clustering threshold is sharp and determine its exact location. We prove that the set of solutions has large conductance below this threshold and that each of the clusters has large conductance above the same threshold.

Our proof constructs a very sparse basis for the set of solutions (or the subset within a cluster). This construction is intimately tied to the construction of specific subgraphs of the hypergraph associated with an instance of k-XORSAT. In order to study such subgraphs, we establish novel local weak convergence results for them.

1. Introduction. An instance of XOR-satisfiability (XORSAT) is specified by an integer n (the number of variables) and by a set of m clauses of the form $x_{i_a(1)} \oplus \cdots \oplus x_{i_a(k)} = b_a$ for $a \in [m] \equiv \{1, \ldots, m\}$. Here, $\oplus$ denotes modulo-2 sum, $\underline{b} = (b_1, \ldots, b_m)$ is a Boolean vector, $b_a \in \{0, 1\}$, specified by the problem instances, and $\underline{x} = (x_1, \ldots, x_n)$ is a vector of Boolean variables $x_i \in \{0, 1\}$ that must be chosen to satisfy the clauses.

Standard linear algebra methods allow us to determine whether a given XOR-SAT instance admits a solution, to find a solution, and even to count the number of solutions, all in polynomial time. In this paper, we shall be interested in the

structural properties of the set of solutions $\mathcal{S} \subseteq \{0, 1\}^n$ of a random k-XORSAT formula. More explicitly, we consider a random XORSAT instance $\mathcal{I}$ that is drawn uniformly at random within the set $\mathbb{G}(n, k, m)$ of instances with m clauses over n variables, whereby each clause involves exactly k variables. The set of solutions $\mathcal{S} = \mathcal{S}(\mathcal{I})$ is then defined as the set of binary vectors $\underline{x}$ that satisfy all m clauses.

Since $\mathcal{I}$ is a random formula, $\mathcal{S}$ is a random subset of the Hamming hypercube. The structural properties of $\mathcal{S}$ are of interest for several reasons. First of all, linear systems over finite fields are combinatorial objects that emerge naturally in a number of fields. Dietzfelbinger and collaborators [18] use a mapping between XORSAT and the matching problem to establish tight thresholds for the performances of Cuckoo Hashing, an archetypal load balancing scheme. Such thresholds are computed by determining thresholds above which the set of solutions $\mathcal{S}$ of a random XORSAT formula becomes empty. The existence of solutions is in turn related to the existence of an even-degree subgraph in a random hypergraph. Random sparse linear systems over finite fields are used to construct capacity achieving error correcting codes [27, 28, 36]. The decodability of such codes is related to the emergence of a nontrivial 2-core in the same random hypergraph—a phenomenon that will play a crucial role in the following. Finally, structured linear systems over finite fields are generated by popular factoring algorithms [25].

In the present paper, we are also motivated by the close analogy between random k-XORSAT and other random ensembles of constraint satisfaction problems (CSPs). The prototypical example of this family is random k-satisfiability (k-SAT). The random k-SAT ensemble can be described in complete analogy to random k-XORSAT with the modification of replacing exclusive OR clauses by OR clauses among variables or their negations. Namely, in k-SAT each clause takes the form $(x'_{i_a(1)} \vee \cdots \vee x'_{i_a(k)})$, whereby $x'_{i_a(\ell)} = x_{i_a(\ell)}$ or $x'_{i_a(\ell)} = \overline{x}_{i_a(\ell)}$. An extensive literature [1, 3, 26, 29, 30, 33] provides strong support for the existence of two sharp thresholds in random k-SAT, as the number of clauses per variable $\alpha = m/n$ grows. First, as α crosses a "satisfiability threshold" $\alpha_{\mathrm{s}}(k)$, random k-SAT formulae pass from being with high probability (w.h.p., i.e., with probability converging to 1 as $n \to \infty$) satisfiable [for $\alpha < \alpha_{\mathrm{s}}(k)$] to being w.h.p. unsatisfiable [for $\alpha > \alpha_{\mathrm{s}}(k)$]. For any $\alpha < \alpha_{\mathrm{s}}(k)$ the set of solutions is therefore nonempty. However, it undergoes a dramatic structural change as α crosses a second threshold $\alpha_{\mathrm{d}}(k) < \alpha_{\mathrm{s}}(k)$. While for $\alpha < \alpha_{\mathrm{d}}(k)$, $\mathcal{S}$ is w.h.p. "well connected" (more precise definitions will be given below), for $\alpha \in [\alpha_{\mathrm{d}}(k), \alpha_{\mathrm{s}}(k)]$ it shatters into an exponential number of clusters. It has been argued that such a "clustered" structure of the space of solutions can have an intimate relation with the failure of standard polynomial time algorithms when applied to random formulae in this regime. The same scenario is thought to hold for a number of random constraint satisfaction problems (including, e.g., proper coloring of random graphs, bicoloring random hypergraphs, Not All Equal-SAT, etc.).

Unfortunately, this fascinating picture is so far only conjectural. Even the best understood element, namely the existence of a satisfiability threshold $\alpha_{\mathrm{s}}(k)$ has

not been established (with the exception of the special case $k = 2$). In an early breakthrough, Friedgut [21] used Fourier-analytic methods to prove the existence of a—possibly n-dependent— sequence of thresholds $\alpha_s(k; n)$. Proving that in fact this sequence can be taken to be n-independent is one of the most challenging open problems in probabilistic combinatorics and random graph theory. Understanding the precise connection between clustering of the space of solutions and computational complexity is an even more daunting task.

Given such outstanding challenges, a fruitful line of research has pursued the analysis of somewhat simpler models. A very interesting possibility is to study k-SAT formulae for large but still bounded values of k. As explained in [1], each SAT clause eliminates only one binary assignment of its k variables, out of 2^k possible assignments of the same variables. Hence, for k large, a single clause has a small effect on the set of solutions, and most binary vectors are satisfying unless the formula includes about 2^k clauses per variable. This results in an "averaging" effect and suitable moment methods provide asymptotically sharp results for large k. In particular, Achlioptas and Peres [4] proved upper and lower bounds on $\alpha_s(k)$ that become asymptotically equivalent (i.e., whose ratio converges to 1) as k gets large. Achlioptas and Coja-Oghlan [1, 2], proved that clustering indeed takes place in an interval of values of α below the satisfiability threshold and obtained upper and lower bounds on the corresponding threshold $\alpha_d(k)$ that are asymptotically equivalent for large k. Finally, Coja-Oghlan [13] proved that solutions can be found w.h.p. in polynomial time for any $\alpha < \alpha_{d,alg}(k)$, whereby $\alpha_{d,alg}(k)$ is asymptotically equivalent to $\alpha_d(k)$ for large k. Intriguingly, no algorithm is known that can provably find solutions in polynomial time for $\alpha \in ((1 + \delta)\alpha_d(k), \alpha_s(k))$, for any $\delta > 0$, and all $k \geq 3$.

XORSAT is a very different example on which rigorous mathematical analysis proved possible, thus providing precious complementary insights. The key simplification is that the set of solutions $\mathcal{S}$ is, in this case, an affine subspace of the Hamming hypercube $\{0, 1\}^n$ (viewed as a vector space over $G\mathbb{F}[2]$). This implies a high degree of symmetry that can be exploited to obtain very sharp characterizations for large n, and any k (we assume throughout that $k \geq 3$, since 2-XORSAT is significantly simpler).

It was proved in [20] that, for $k = 3$, there exists an n-independent threshold $\alpha_s(k)$ such that a random k-XORSAT instance is w.h.p. satisfiable if $\alpha < \alpha_s(k)$ and unsatisfiable if $\alpha > \alpha_s(k)$. The proof constructs a subformula, by considering the 2-core of the hypergraph associated with the XORSAT instance. One can then prove that the original formula is satisfiable if and only if the 2-core subformula is. The threshold for the latter can be determined exactly using the second moment method. The proof was extended to all $k \geq 4$ in [18].

The existence of a 2-core in a random XORSAT formula has a sharp threshold when the number of clauses per variable α crosses a value $\alpha_{core}(k)$. This was argued to be intimately related to the appearance of clusters. In particular, [12, 31]

give an argument[3] showing that, above $\alpha_{\mathrm{core}}(k)$, the space of solutions shatters into exponentially many clusters. In other words, $\alpha_{\mathrm{core}}(k)$ is an upper bound on the clustering threshold. [31] further shows that, for $\alpha < \alpha_{\mathrm{core}}(k)$, a particular coordinate of a solution can be changed by changing $O(1)$ other variables on average, without leaving the space of solutions. If this argument is pushed a step further, one can show that, w.h.p., any coordinate can be changed by flipping at most $O(\log n)$ other coordinates. This suggests that it *may* be possible to concatenate a sequence of such flips to connect any two solutions via a path through the solution space, with $O(\log n)$ steps. However, the analysis [31] does not imply that this is the case, as it does not address the main challenge, namely to construct a path from any solution to any other solution. In this work, we solve this problem and provide the first proof of a lower bound of $\alpha_{\mathrm{core}}(k)$ on the clustering threshold α_{d}, thus establishing that indeed $\alpha_{\mathrm{d}}(k) = \alpha_{\mathrm{core}}(k)$. For $\alpha > \alpha_{\mathrm{d}}(k)$ we prove a sharp characterization of the decomposition into clusters.

As mentioned above, random k-XORSAT formulae can be solved in polynomial time using linear algebra methods, and this appears to be insensitive to the clustering threshold. Nevertheless, an intriguing algorithmic phase transition might take place *exactly* at the clustering threshold $\alpha_{\mathrm{d}}(k)$. For any $\alpha < \alpha_{\mathrm{d}}(k)$, solutions can be found in time linear in the number of variables (the algorithm is in fact an important component of our proof). On the other hand, no algorithm is known that finds a solution in linear time for $\alpha \in (\alpha_{\mathrm{d}}(k), \alpha_{\mathrm{s}}(k))$. We think that our proof sheds some light on this phenomenon.

1.1. *Main result.*   In this paper, we obtain two sharp results characterizing the clustering phase transition for random k-XORSAT:

(i)   We exactly determine the clustering threshold $\alpha_{\mathrm{d}}(k)$, proving that the space of solutions is w.h.p. well connected for $\alpha < \alpha_{\mathrm{d}}(k)$, and instead shatters into exponentially many clusters for $\alpha \in (\alpha_{\mathrm{d}}(k), \alpha_{\mathrm{s}}(k))$.

(ii)   We determine the exponential growth rate of the number of clusters, that is, we show that this is w.h.p. $\exp\{n\Sigma(\alpha; k) + o(n)\}$ where $\Sigma(\alpha; k)$ is a nonrandom function which is explicitly given. We prove that each of the clusters is itself "well connected."

This is therefore the first random CSP ensemble for which a sharp threshold for clustering is proved.

Earlier literature fell short of establishing (i) since it did not provide any argument for connectedness below $\alpha_{\mathrm{d}}(k)$. Also, informal calculations only suggested a lower bound on the number of clusters, but did not establish (ii) since they did not prove connectedness of each cluster by itself. The situation is akin to the analysis

---

[3]The argument of [12, 31] is essentially rigorous, but does not deal with several technical steps.

of Markov Chain Monte Carlo methods: It is often significantly more challenging to prove rapid mixing (connectedness of the space of configurations) than the opposite (i.e., to find bottlenecks).

One important novelty is that the notion of connectedness used here is very strong and goes beyond path connectivity, which was used earlier for k-SAT [1, 2]. We use a properly defined notion of *conductance* which we think can be applied to a broader set of CSPs, and has the advantage of being closely related to important algorithmic notions (fast mixing for MCMC and expansion). Given a subset of the hypercube $\mathcal{S} \subseteq \{0, 1\}^n$, and a positive integer ℓ, we define the conductance of $\mathcal{S}$ as follows. Construct the graph $\mathcal{G}(\mathcal{S}, \ell)$ with vertex set $\mathcal{S}$ and an edge connecting $\underline{x}, \underline{x}' \in \mathcal{S}$ if and only if $d(\underline{x}, \underline{x}') \leq \ell$ [here and below, $d(\cdot, \cdot)$ denotes the Hamming distance, i.e., $d(\underline{x}, \underline{x}') = |\{i : 1 \leq i \leq n, x_i \neq x_i'\}|$, where $\underline{x} = (x_1, x_2, \ldots, x_n)$ and similarly for $\underline{x}'$, and $|B|$ denotes the cardinality of the set B]. Then we define the *ℓth conductance of $\mathcal{S}$* as the graph conductance of $\mathcal{G}(\mathcal{S}, \ell)$, namely

$$(1) \qquad \Phi(\mathcal{S}; \ell) \equiv \min_{A \subseteq \mathcal{S}} \frac{\mathrm{cut}_{\mathcal{G}(\mathcal{S}, \ell)}(A, \mathcal{S} \setminus A)}{\min(|A|, |\mathcal{S} \setminus A|)},$$

where, for a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, and any $B \subseteq \mathcal{V}$, we define

$$\mathrm{cut}_{\mathcal{G}}(B, \mathcal{V} \setminus B) \equiv \big|\{e \in \mathcal{E} : \text{Exactly one of the two endpoints of } e \text{ is in } B\}\big|.$$

Notice that we measure the volume of a set by the number of its vertices instead of the sum of its degrees.[4]

We define the distance between two subsets of the hypercube $\mathcal{S}_1, \mathcal{S}_2 \subseteq \{0, 1\}^n$ as

$$d(\mathcal{S}_1, \mathcal{S}_2) \equiv \min_{\underline{x} \in \mathcal{S}_1, \underline{x}' \in \mathcal{S}_2} d(\underline{x}, \underline{x}').$$

For our statements, $k \geq 3$ is always fixed, together with a sequence $m(n) = \alpha n$.

We say that a sequence of events $(\mathsf{E}_n)_{n>0}$ occurs *with high probability* (w.h.p.) if $\lim_{n \to \infty} \mathbb{P}(\mathsf{E}_n) = 1$. (We refer to Section 2 for a formal definition of the underlying probability space.)

THEOREM 1.   *Let $\mathcal{S}$ be the set of solutions of a random k-XORSAT formula with n variables and $m = n\alpha$ clauses. For any $k \geq 3$, let $\alpha_{\mathrm{d}}(k)$ be defined as*

$$(2) \qquad \alpha_{\mathrm{d}}(k) \equiv \sup\big\{\alpha \in [0, 1] : z > 1 - e^{-k\alpha z^{k-1}}, \forall z \in (0, 1)\big\}.$$

---

[4]This difference is irrelevant for $\alpha < \alpha_{\mathrm{d}}(k)$ since in this case $\mathcal{S}$ will be taken to be an affine subspace of $\{0, 1\}^n$, and hence $\mathcal{G}(\mathcal{S}, \ell)$ will be a regular graph. For $\alpha \in (\alpha_{\mathrm{d}}(k), \alpha_{\mathrm{s}}(k))$, $\mathcal{S}$ will be constructed as the union of a "small" number of affine spaces, and hence $\mathcal{G}(\mathcal{S}, \ell)$ should be approximately regular. We keep the definition (1) since it simplifies our statements.

1. *If $\alpha < \alpha_d(k)$, there exists $C = C(\alpha, k) < \infty$ such that, w.h.p., $\Phi(\mathcal{S}; (\log n)^C) \geq 1/2$.*

2. *If $\alpha \in (\alpha_d(k), \alpha_s(k))$, then there exists $\varepsilon = \varepsilon(k; \alpha) > 0$ such that, w.h.p., $\Phi(\mathcal{S}; n\varepsilon) = 0$.*

3. *If $\alpha \in (\alpha_d(k), \alpha_s(k))$, and $\delta > 0$ is arbitrary, then there exist constants $C = C(\alpha, k) < \infty$, $\varepsilon = \varepsilon(\alpha, k) > 0$, $\Sigma = \Sigma(\alpha, k) > 0$, and a partition of the set of solutions $\mathcal{S} = \mathcal{S}_1 \cup \cdots \cup \mathcal{S}_N$, such that, w.h.p., the following properties hold:*
   (a) *For each $a \in [N]$, we have $\Phi(\mathcal{S}_a; (\log n)^C) \geq 1/2$.*
   (b) *For each $a \neq b \in [N]$, we have $d(\mathcal{S}_a, \mathcal{S}_b) \geq n\varepsilon$.*
   (c) *$\exp\{n(\Sigma - \delta)\} \leq N \leq \exp\{n(\Sigma + \delta)\}$. Further, letting Q be the largest positive solution of $Q = 1 - \exp\{-k\alpha Q^{k-1}\}$ and $\widehat{Q} \equiv Q^{k-1}$, we have $\Sigma(\alpha, k) = Q - k\alpha\widehat{Q} + (k-1)\alpha Q\widehat{Q}$.*

1.2. *Conductance and sparse basis.*   We will prove Theorem 1 by obtaining a fairly complete description of the set $\mathcal{S}$ both above and below $\alpha_d(k)$. In a nutshell, for $\alpha < \alpha_d(k)$, $\mathcal{S}$ admits a sparse basis, while for $\alpha > \alpha_d(k)$ each of the clusters $\mathcal{S}_1, \ldots, \mathcal{S}_N$ admits a sparse basis but their union does not. This is particularly suggestive of the connection between the clustering phase transitions and algorithm performance. Below $\alpha_d(k)$ the space of solutions admits a succinct explicit representation [in $O(n(\log n)^C)$ bits]. Above $\alpha_d(k)$, we can produce a representation that is succinct but implicit (as solutions of a given formula), or explicit but prolix [no basis is known that can be encoded in $o(n^2)$ bits].

Given a linear subspace $\mathcal{S} \subseteq \{0, 1\}^n$, we say that it admits an s-sparse basis if there exist vectors $\underline{x}^{(l)} \in \mathcal{S}$ for $l \in \{1, \ldots, D\}$ such that $d(\underline{x}^{(l)}, \underline{0}) \leq s$ and $(\underline{x}^{(l)})_{l=0}^D$ form a basis for $\mathcal{S}$. The latter means that the vectors are linearly independent and $\mathcal{S} = \{\sum_{l=1}^D a_l \underline{x}^{(l)} : (a_l)_{l=0}^D \in \{0, 1\}^D\}$.

We say that an affine space $\mathcal{S} \subseteq \{0, 1\}^n$ admits an s-sparse basis if, for $\underline{x}^{(0)} \in \mathcal{S}$, the linear subspace $\mathcal{S} - \underline{x}^{(0)}$ admits an s-sparse basis. The property of having a sparse basis indeed implies large conductance. The proof is immediate.

LEMMA 1.1.   *If the affine subspace $\mathcal{S} \subseteq \{0, 1\}^n$ admits a s-sparse basis, then $\Phi(\mathcal{S}; s) \geq 1/2$.*
Vice versa, assume that $\Phi(\mathcal{S}; s) = 0$. Then $\mathcal{S}$ does not admit a s-sparse basis.

PROOF.   We can assume, without loss of generality, that $\mathcal{S}$ is a linear space. Let d be its dimension. Further, given a graph $\mathcal{G}$, let, with a slight abuse of notation

$$(3) \qquad \Phi(\mathcal{G}) \equiv \min_{A \subseteq \mathcal{S}} \frac{\mathrm{cut}_{\mathcal{G}}(A, \mathcal{S} \setminus A)}{\min(|A|, |\mathcal{S} \setminus A|)},$$

so that $\Phi(\mathcal{S}; \ell) = \Phi(\mathcal{G}(\mathcal{S}; \ell))$.

Assume that $\mathcal{S}$ admits a s-sparse basis. This immediately implies the graph $\mathcal{G}(\mathcal{S}, s)$ contains a spanning subgraph that is isomorphic to the d-dimensional hypercube $\mathcal{H}_d$. Further, $\mathcal{G} \mapsto \Phi(\mathcal{G})$ is monotone increasing in the edge set of $\mathcal{G}$.

Therefore, $\Phi(\mathcal{S}; s) \geq \Phi(\mathcal{H}_d) \geq 1/2$ where the last inequality follows from the standard isoperimetric inequality on the hypercube [23]. $\square$

The characterization of the solution space in terms of sparsity of its basis is given below.

THEOREM 2.   *Let $\mathcal{S}$ be the set of solutions of a random k-XORSAT formula with n variables and $m = n\alpha$ clauses. For any $k \geq 3$, let $\alpha_d(k)$ be defined as per equation* (2). *Then the following hold*:

1. *If $\alpha < \alpha_d(k)$, there exists $C = C(\alpha, k) < \infty$ such that, w.h.p., $\mathcal{S}$ admits a $(\log n)^C$-sparse basis.*
2. *If $\alpha \in (\alpha_d(k), \alpha_s(k))$, and $\delta > 0$ is arbitrary, then there exist constants $C = C(\alpha, k) < \infty$, $\varepsilon = \varepsilon(\alpha, k) > 0$, $\Sigma = \Sigma(\alpha, k) > 0$, and a partition of the set of solutions $\mathcal{S} = \mathcal{S}_1 \cup \cdots \cup \mathcal{S}_N$, such that, w.h.p., the following properties hold*:
   (a) *For each $a \in [N]$, $\mathcal{S}_a$ admits a $(\log n)^C$-sparse basis.*
   (b) *For each $a \neq b \in [N]$ we have $d(\mathcal{S}_a, \mathcal{S}_b) \geq n\varepsilon$.*
   (c) $\exp\{n(\Sigma - \delta)\} \leq N \leq \exp\{n(\Sigma + \delta)\}$. *Further, Σ is given by the same expression given in Theorem* 1.

Clearly, this theorem immediately implies Theorem 1 by applying Lemma 1.1. The rest of this paper is devoted to the proof of Theorem 2.

1.3. *Further technical contributions.*   To a given a XORSAT instance $\mathcal{I}$, we can associate a bipartite graph ("factor graph") with vertex sets F (*factor* or *check* nodes) corresponding to equations, and V (*variable* nodes) variables. The edge set E includes those pairs $(a, i) \in F \times V$ such that variable x_i participates in the ath equation. The construction of the sparse basis in Theorem 2 relies heavily on a characterization of the random factor graph associated to a random XORSAT instance. This could be gleaned from the proof of [18, 20] that construct the 2-core of G. In order to prove Theorem 2, we characterize a larger subgraph that we refer to as the *backbone* of G. This subgraph has the following interpretation: if two solutions $\underline{x}$ and $\underline{x}'$ coincide on the core, then they coincide on every vertex of the backbone.

The 2-core of the random graph G was studied in a number of papers [14, 27, 32, 35]. The key tool in these works is the analysis of an iterative procedure that constructs the 2-core in $\Theta(n)$ iterations. This procedure has an important property: At each step, the resulting graph remains uniformly random, given a small number of parameters (essentially, its degree distribution). Thanks to this property, the analysis of [14, 27, 32, 35] is reduced to the study of a Markov chain in $\mathbb{Z}^2$. This is done by showing that sample paths of this chain are shown to concentrate around solutions of a certain ordinary differential equation.

Our analysis of the backbone has a similar starting point, namely the study of an iterative procedure that constructs the backbone (indeed we define formally the

backbone as the fixed point of this procedure). Unfortunately, the graphs generated by this procedure are not uniformly random, conditional on a small number of parameters. Hence, the techniques [14, 27, 32, 35] do not apply. We overcome this difficulty by characterizing the large-n limit of its fixed point using the theory of local weak convergence. This is in turn challenging because the fixed point is not, *a priori*, a local function of G.

We consider this characterization of the backbone, and its proof, to be a contribution of independent interest.

For describing the iterative procedure, we use the language of message passing algorithms, and will refer to it as to "belief propagation" (BP), as the same algorithm is also of interest in iterative coding; see [29, 36]. Given a factor graph $G = (F, V, E)$, the algorithm updates $2|E|$ *messages* indexed by directed edges in G. In other words, for each $(a, v) \in E$, $a \in F$ and $v \in V$, and any iteration number $t \in \mathbb{N}$, we have two messages $v_{v \to a}^t$, and $\widehat{v}_{a \to v}^t$, taking values in $\{0, *\}$. For $t \geq 1$, messages are computed following the update rules:

$$
(4) \qquad v_{v \to a}^t = \begin{cases} *, & \text{if } \widehat{v}_{b \to v}^{t-1} = * \text{ for all } b \in \partial v \setminus a, \\ 0, & \text{otherwise,} \end{cases}
$$

and

$$
(5) \qquad \widehat{v}_{a \to v}^t = \begin{cases} 0, & \text{if } v_{u \to a}^t = 0 \text{ for all } u \in \partial a \setminus v, \\ *, & \text{otherwise.} \end{cases}
$$

We call this algorithm BP$_0$ when all messages are initialized to 0: $v_{v \to a}^0 = \widehat{v}_{a \to v}^0 = 0$ for all $(a, v) \in E$. It is not hard to see that BP$_0$ is monotone,[5] in the sense that messages only change from 0 to $*$, and hence converges to a fixed point $v_{v \to a}^*$.

It is easy to check (see Lemma 4.3 below) that the core of G coincides with the subgraph induced by the factor nodes that receive no $*$ message at the fixed point of BP$_0$. The backbone is instead the subgraph induced by factor nodes that receive at most one $*$ message at the fixed point.

Denote by $\tilde{\mu}_n^*$ the probability distribution on rooted factor graphs with marks on the edges constructed as follows. Draw a graph uniformly at random from $\mathbb{G}(n, k, m)$. Choose a uniformly random variable node $i \in V$ as the root. Mark the edges (in each direction) with the messages corresponding to the BP$_0$ fixed point $v_{v \to a}^*$.

We next construct a random tree $\widetilde{\mathcal{T}}_*(\alpha, k)$ with marks on the directed edges as follows. Marks take values in $\{0, *\}$ and to each undirected edge we associate a mark for each of the two directions. We will refer to the direction toward the root as to the "upward" direction, and to the opposite one as to the

---

[5]This can be established by induction: Since we start with all 0s, clearly messages can only change from 0 to $*$ in the first iteration. Thereafter, this holds inductively for each subsequent iteration since each of the update rules is monotone in the sense that if the incoming messages only change from 0 to $*$, then the same holds for the outgoing messages.

"downward" direction. The marks correspond to fixed-point BP messages, and we will call them messages as well in what follows. First, consider only edges directed upward. This is a multitype Galton–Watson (GW) tree. At the root generate Poisson($k\alpha$) offsprings, and mark each of the edges to 0 independently with probability $\widehat{Q}$, and to $*$ otherwise. At a nonroot variable node, if the parent edge is marked 0, generate Poisson($k\alpha(1 - \widehat{Q})$) descendant edges marked $*$ and Poisson$_{\geq 1}(k\alpha\widehat{Q})$ descendant edges marked 0 [here Poisson$_{\mathsf{E}}(\lambda)$ denotes a Poisson random variable with parameter λ conditional on E]. If the parent edge is marked $*$, generate Poisson($k\alpha(1 - \widehat{Q})$) descendant edges marked $*$ and no descendant edges marked 0. At a factor node, if the parent edge is marked 0, generate $k - 1$ descendant edges marked 0. If the parent node is marked $*$, generate $M \sim \text{Binom}_{\leq k-2}(k - 1, Q)$ descendants marked 0, and $k - 1 - M$ descendants marked $*$.

For edges directed downward, marks are generated recursively following the usual BP rules, cf. equations (4), (5), starting from the top to the bottom. It is easy to check that with this construction, the marks in $\widetilde{\mathcal{T}}_*(\alpha, k)$ correspond to a BP fixed point. Given a factor graph $G = (F, V, E)$, we use $\mathsf{B}_G(v, t)$ to denote the ball of radius t centered at node $v \in V$. This ball is defined inductively as follows: The $\mathsf{B}_G(v, 0)$ consists of node v alone and no edges. For $t > 0$, the $\mathsf{B}_G(v, t)$ includes $\mathsf{B}_G(v, t - 1)$. In addition, it includes all factor nodes connected to variable nodes in $\mathsf{B}_G(v, t - 1)$ and associated edges, and all variable nodes connected to those factor nodes and associated edges. [Thus, $\mathsf{B}_G(v, t)$ includes nodes and edges up to a distance t from v, where variable nodes are said to be separated by distance 1 if they are connected to the same factor node.]

DEFINITION 1.2.   Let $\{G_n\}$, $G_n = (F_n, V_n, E_n)$ be a sequence of (random) factor graphs. Let $\mu_n^{(t)}$ denote the empirical probability distribution of $\mathsf{B}_{G_n}(v, t)$ when $v \in V_n$ is uniformly random. Explicitly, for any locally finite rooted graph $\mathcal{T}_0$ of depth at most t,

$$(6) \qquad \mu_n^{(t)} \equiv \frac{1}{n} \sum_{v \in V_n} \mathbb{I}\big(\mathsf{B}_{G_n}(v, t) \simeq \mathcal{T}_0\big),$$

(with $\simeq$ denoting equality up to graph vertex relabeling.) We say that $\{G_n\}$ *converges locally almost surely* to the measure μ on rooted graphs if, for any finite t, and any locally finite rooted graph $\mathcal{T}_0$ of depth at most t, we have

$$(7) \qquad \lim_{n \to \infty} \mu_n^{(t)}(\mathcal{T}_0) = \mu^{(t)}(\mathcal{T}_0)$$

holds almost surely with respect to the graph law. Here, $\mu^{(t)}$ denotes the marginal of μ with respect to a ball of radius t around the root.

The same notion of local graph convergence was used earlier in the literature, for instance, in [15–17]. Given a random graph distribution, we first draw a sequence

of $\{G_n\}_{n \geq 1}$, and then check that $\mu_n^{(t)}(\mathcal{T}_0) \to \mu^{(t)}(\mathcal{T}_0)$ with probability one. It is worth emphasizing the difference from a weaker notion (that we never use below), whereby we only check $\mathbb{E}_{G_n} \mu_n^{(t)}(\mathcal{T}_0) \to \mu^{(t)}(\mathcal{T}_0)$, with $\mathbb{E}_{G_n}$ denoting expectation with respect to the graph distribution. In particular, establishing almost sure local graph convergence is more challenging that proving convergence of the expectation $\mathbb{E}_{G_n} \mu_n^{(t)}(\mathcal{T}_0)$ since it requires to control the deviations of the subgraph counts $\mu_n^{(t)}(\mathcal{T}_0)$. With this clarification, we shall occasionally drop the "almost surely" in "converges locally almost surely."

As part of our proof of Theorem 2, we obtain the following result, which may be of independent interest. (We refer to the next section for a complete definition of the underlying probability space.)

THEOREM 3. *The sequence $\{\tilde{\mu}_n^*\}_{n \geq 0}$ converges locally almost surely to the probability distribution of $\widetilde{\mathcal{T}}_*$.*

Theorem 3 is proved in Section 4.
Besides this, our proof uses several other ideas:

- We show that Theorem 3 can be used to extend the low weight core solutions to low weight solutions of the whole XORSAT instance (see Section 8).
- We show that the periphery (the complement of the core in G) is uniformly random with a given degree sequence, conditioned on being "peelable." We estimate precisely this degree distribution, and show that the periphery is indeed peelable with positive probability for that degree sequence (see Section 6).
- In addition to the fixed point characterization, we obtain a precise characterization of the convergence rate of BP_0 (see Section 4), which allows us to bound the sparsity of the basis constructed.
- For $\alpha > \alpha_d$, convergence to the BP fixed point is geometric rather than quadratic. In this regime, we show that in fact there are "strings" of degree 2 variable nodes that slow down convergence but do not prevent the construction of a sparse basis. We bound the sparsity by defining a certain "collapse" operator on such strings (see Section 5).

1.4. *Outline of the paper.* In Section 2, we define some basic concepts and notation. Section 3 describes the construction of clusters and sparse bases, and uses this construction to prove Theorem 2. Several basic lemmas necessary for the proof are stated in this section.

Section 4 introduces a certain *belief propagation* (BP) algorithm and a technical tool called *density evolution*, that play a key role in our analysis: The BP algorithm naturally decomposes the linear system into a "backbone" (consisting roughly of the 2-core and the variables implied by it) and a "periphery." Density evolution allows us to track the progress of BP, eventually facilitating a tight characterization of basic parameters (like number of nodes) of the backbone and periphery.

Section 5 bounds the number of iterations of a "peeling" algorithm (related to BP) that plays a key role in our construction of a sparse basis. Section 6 proves a sharp characterization of the periphery. Together, this yields the first (large) set of basis vectors.

Section 7 shows the 2-core has very few sparse solutions, leading to well separated, small, "core-clusters." Section 8 shows how to produce a sparse solution of the linear system corresponding to each sparse solution of the 2-core subsystem. This yields the second (small) set of basis vectors in our construction.

Several technical lemmas are deferred to the Appendices.

A short version of this paper was presented at the ACM-SIAM Symposium on Discrete Algorithms SODA 2012.

2. Random k-XORSAT: Definitions and notation. As described in the Introduction, each k-XORSAT clause is actually a linear equation over $\mathsf{GF}[2]$: $x_{i_a(1)} \oplus \cdots \oplus x_{i_a(k)} = b_a$, for $a \in [m] \equiv \{1, \ldots, m\}$. Introducing a vector $\underline{h}_a \in \{0, 1\}^n$, with nonzero entries only at positions $i_1(a), \ldots, i_k(a)$, this can be written as $\underline{h}_a^T \underline{x} = b_a$. Hence, an instance is completely specified by the pair $(\mathbb{H}, \underline{b})$ where $\mathbb{H} \in \{0, 1\}^{m \times n}$ is a matrix with rows $\underline{h}_1^T, \ldots, \underline{h}_m^T$ and $\underline{b} = (b_1, \ldots, b_m)^T \in \{0, 1\}^m$. The space of solutions is therefore $\mathcal{S} \equiv \{\underline{x} \in \{0, 1\}^n : \mathbb{H}\underline{x} = \underline{b} \bmod 2\}$. If $\mathcal{S}$ has at least one element $\underline{x}^{(0)}$, then $\mathcal{S} \oplus \underline{x}^{(0)}$ is just the set of solutions of the homogeneous linear system corresponding to $\underline{b} = \underline{0}$ (the kernel of $\mathbb{H}$). In the following we shall always assume $\alpha < \alpha_s(k)$, so that $\mathcal{S}$ is nonempty w.h.p. [18]. Note that, if $\mathcal{S}$ is nonempty, then $\mathcal{S} = \mathcal{S}_0 \oplus \underline{x}_0$ where $\underline{x}_0 \in \mathcal{S}$ is any solution of the original system and $\mathcal{S}_0$ is the set of solutions of the homogeneous linear system $\mathbb{H}\underline{x} = \underline{0}$. Since we are only interested in geometric properties of the set of solutions that are invariant under translation, we will assume hereafter that $\underline{b} = \underline{0}$, and hence $\mathcal{S} = \mathcal{S}_0$.

An XORSAT instance is therefore completely specified by a binary matrix $\mathbb{H}$, or equivalently by the corresponding factor graph $G = (F, V, E)$. This is a bipartite graph with two sets of nodes: F (*factor* or *check* nodes) corresponding to rows of $\mathbb{H}$, and V (*variable* nodes) corresponding to columns of $\mathbb{H}$. The edge set E includes those pairs (a, i), $a \in F$, $i \in V$ such that $\mathbb{H}_{ai} = 1$. We denote by $\mathbb{G}(n, k, m)$ the set of all factor graphs with n labeled variable nodes and m labeled check nodes, each having degree exactly k (with no double edges). Note that $|\mathbb{G}(n, k, m)| = \binom{n}{k}^m$. With a slight abuse of notation, we will denote by $\mathbb{G}(n, k, m)$ also the uniform distribution over this set, and write $G \sim \mathbb{G}(n, k, m)$ for a uniformly random such graph.

For $v \in V$ or $v \in F$, we denote by $\deg_G(v)$, the degree of node v in graph G (omitting the subscript when clear from the context) and we let ∂v denote the set of neighbors of v. We define the distance with respect to G between two variable nodes $i, j \in V$, denoted by $d_G(i, j)$ as the length of the shortest path from i to j in G, whereby the length of a path is the number of check nodes encountered along the path. Given a vector $\underline{x}$, we denote by $\underline{x}_A = (x_i)_{i \in A}$ its restriction to A. The cardinality of set A is denoted by $|A|$.

We only consider the "interesting" case $k \geq 3$, and the asymptotics $m, n \to \infty$ with $m/n \to \alpha$ and $\alpha \in [0, \alpha_s(k))$, where $\alpha_s(k)$ is the satisfiability threshold. Hence, $\mathbb{H}$ has w.h.p. maximum rank, that is, $\mathrm{rank}(\mathbb{H}) = m$ [29].

DEFINITION 2.1.   Let $F_0 \subseteq F$. The *subgraph induced* by F_0 is defined as (F_0, V_0, E_0) where $V_0 \equiv \{i \in V : \partial i \cap F_0 \neq \varnothing\}$ and $E_0 \equiv \{(a, i) \in E : a \in F_0, i \in V_0\}$. A *check-induced* subgraph is the subgraph (F_0, V_0, E_0) induced by some $F_0 \subseteq F$. Similarly, we can define the subgraph induced by $V_0 \subseteq V$, and *variable-induced* subgraphs.

Let $F_0 \subseteq F$, $V_0 \subseteq V$. The *subgraph induced* by (F_0, V_0) is defined as (F_0, V_0, E_0) where $E_0 \equiv \{(a, i) \in E : a \in F_0, i \in V_0\}$.

DEFINITION 2.2.   A *stopping set* is a check-induced subgraph with the property that every variable node has degree larger than one with respect to the subgraph. The 2-*core* of G is its maximal stopping set.

Notice that the *maximal* stopping set of G is uniquely defined because the union of two stopping sets is a stopping set.

All of our statements are with respect to the following probability space, for a fixed $k \geq 3$, and an integer sequence $\{m(n)\}_{n \in \mathbb{N}}$. For each n, we let $m = m(n)$ and consider the finite set $\Omega_n = \mathbb{G}(n, k, m)$ of k-XORSAT instances with n variables and m clauses. Formally, each element of $\mathbb{G}(n, k, m)$ is given by a pair $(\mathbb{H}, \underline{b})$ where $\mathbb{H} \in \{0, 1\}^{m \times n}$ is a matrix with k nonzero elements per row and $\underline{b} \in \{0, 1\}^m$. [In the proofs, we shall occasionally replace $\mathbb{G}(n, k, m)$ by slightly different sets—defined therein—for technical convenience. The connection will be made clear.]

Since Ω_m is finite, it is straightforward to endow it with the uniform probability measure $\mathbb{P}_n$ over the complete σ-algebra 2^{Ω_n}. The probability space underlying all of our statements is the product space $\Omega = \bigtimes_{n \in \mathbb{N}} \Omega_n$, with product probability measure $\mathbb{P} = \bigtimes_{n \in \mathbb{N}} \mathbb{P}_n$. An event $\mathsf{E} \subseteq \Omega$ is a an element of the product σ-algebra. As a special example, let $f_n : \Omega_n \to \mathbb{R}$ be a sequence of functions, and $\omega = (\omega_i)_{i \in \mathbb{N}} \in \Omega$. Then existence of the limit $\lim_{n \to \infty} f_n(\omega_n)$ is a well defined event in Ω.

With a slight abuse of language, we will identify any set $\mathsf{E}_n \subseteq \Omega_n$ with an event, namely with the cylindrical set $C(\mathsf{E}_n) \equiv \{\omega = (\omega_i)_{i \in \mathbb{N}} \in \bigtimes_{i \in \mathbb{N}} \Omega_i : \omega_n \in \mathsf{E}_n\}$. We will typically write E_n for $C(\mathsf{E}_n)$ and $\mathbb{P}(\mathsf{E}_n) = \mathbb{P}(\{\omega_n \in \mathsf{E}_n\})$ for the probability of such an event. We say that E_n occurs *with high probability* (*w.h.p.*) if $\lim_{n \to \infty} \mathbb{P}(\mathsf{E}_n) = 1$. We say that a sequence of events $(\mathsf{E}_n)_{n > 0}$ occurs *eventually almost surely* if $\lim_{n_0 \to \infty} \mathbb{P}(\bigcap_{n \geq n_0} C(\mathsf{E}_n)) = 1$.

Note that, with this probability space, the notion of local almost sure convergence in Definition 1.2 is well defined. Note that our main results (Theorems 1 and 2) are "with high probability results," and hence do not require the definition of a common probability space for different graph sizes. This is indeed mainly a matter of technical convenience (and is of course needed for Theorem 3).

A key fact to be used in the following is that a giant 2-core appears abruptly at $\alpha_d(k)$. Forms of the following statement appear in [14, 27, 32].

THEOREM 4 ([14, 27, 32]).   *Assume* $\alpha < \alpha_d(k)$. *Then w.h.p.*, *a graph* $G \sim \mathbb{G}(n, k, m)$ *does not contain any stopping set.*
Vice versa, *assume* $\alpha > \alpha_d(k)$. *Then there exists* $C(k) > 0$ *such that*, *w.h.p.*, *a graph* G *drawn uniformly at random from* $\mathbb{G}(n, k, m)$ *contains a 2-core of size larger than* $C(k)n$.

We will often refer to the depth-t neighborhood of a node v in G.

DEFINITION 2.3.   Given a node $v \in V$ and an integer t, let $V' = \{u : u \in V, d_G(u, v) \leq t\}$. Then the *ball of radius* t *around node* v is defined as the (variable-induced) subgraph $\mathsf{B}_G(v, t)$ induced by V'. With an abuse of notation, we will use the same notation for the set of variable nodes in $\mathsf{B}_G(v, t)$. Lastly, we define $|\mathsf{B}_G(v, t)|$ to be the number of variable nodes in the subgraph $\mathsf{B}_G(v, t)$.

We will occasionally work with certain random infinite rooted factor graphs, with marks on the edges or vertices. (Note that a factor graph can be regarded as an ordinary graph, with additional marks on the vertices to distinguish "variable nodes" from "factor nodes.") A useful concept in this context is the one of "unimodular" random rooted graphs, that we briefly recall next. For a more complete presentation, we refer to the overview paper by Aldous and Lyons [5].

Informally, a random rooted (marked) graph is unimodular if it looks the same (in distribution), when the root is moved to any other vertex. In order to formalize this notion, we denote by $\mathcal{G}_*$ the space of locally finite rooted graphs, with marks on the vertices or edges (we assume marks to belong to some fixed finite set for simplicity). We view two graphs that differ by an isomorphism as identical. This space can be endowed by a metric that metrizes local convergence, and hence a Borel σ-algera.

Analogously, we denote by $\mathcal{G}_{**}$ the space of doubly rooted graphs [a *doubly rooted graph* is a graph with two distinguished vertices, i.e., a triple (G, u, v) where $G = (V, E)$ is a graph, and $u, v \in V$]. As for the simply rooted case, $\mathcal{G}_{**}$ can be made into a complete metric space; we regard it as a measurable space endowed with the Borel σ-algebra.

DEFINITION 2.4.   Let $(G, \varnothing)$ be a random rooted graph with root $\varnothing$. We say that $(G, \varnothing)$ is unimodular if, for any measurable function $f : \mathcal{G}_{**} \to \mathbb{R}_{\geq 0}$, $(G, u, v) \mapsto f(G, u, v)$, we have

$$(8) \qquad \mathbb{E}\left[\sum_{v \in V(G)} f(G, \varnothing, v) \right] = \mathbb{E}\left[\sum_{v \in V(G)} f(G, v, \varnothing) \right].$$

Consequences, and equivalent versions of unimodularity can be found in [5, 34].

SYNCHRONOUS PEELING (Graph $G = (F, V, E)$)

$F' \leftarrow F$
$V' \leftarrow V$
$E' \leftarrow E$
$J_0 \leftarrow (F, V, E), t = 0$
While J_t has a variable node of degree ≤ 1 do
    $t \leftarrow t + 1$
    $V_t \leftarrow \{v \in V' : \deg_{G_{t-1}}(v) \leq 1\}$
    $F_t \leftarrow \{a \in F' : (v, a) \in \quad E' \text{ for some } v \in V_t\}$
    $E_t \leftarrow \{(v, a) \in E' : a \in F_t, v \in V'\}$
    $F' \leftarrow F' \setminus F_t$
    $V' \leftarrow V' \setminus V_t$
    $E' \leftarrow E' \setminus E_t$
    $J_t \leftarrow (F', V', E')$
End While
$T_C \leftarrow t$
$G_C \leftarrow G'$
Return $(G_C, T_C, (F_t)_{t=1}^{T_C}, (V_t)_{t=1}^{T_C}, (J_t)_{t=1}^{T_C})$

---

3. Proof of Theorem 2. In this section, we describe the construction of clusters and sparse bases within the clusters [or for the whole space of solutions for $\alpha \in [0, \alpha_d(k))$]. The analysis of this construction is given in Section 3.3 in terms of a few technical lemmas. Finally, the formal proof of Theorem 2 is given in Section 3.4.

3.1. *Construction of the sparse basis.* The construction of a sparse basis, which is at the heart of Theorem 2, is based on the following algorithm, formally stated in Table 1. The algorithm constructs a sequence of residual factor graphs $(J_t)_{t \geq 0}$, starting with the instance under consideration $J_0 = G$. At each step, the new graph is constructed by removing all variable nodes of degree one or zero, their adjacent factor nodes, and all the edges adjacent to these factor nodes. We refer to the algorithm as *synchronous peeling* or simply *peeling*.

We denote the sets of nodes and edges removed at step (or round) $t \geq 1$ by (F_t, V_t, E_t), so that $J_{t-1} = (F_t, V_t, E_t) \cup J_t$. Notice that, at each step, the residual graph J_t is check-induced. The algorithm halts when the residual graph does not contain any variable node of degree smaller than two. We let the total number of iterations be $T_C(G)$, where we will drop the explicit dependence on G when it is clear from context. The final residual graph is then $J_{T_C} \equiv G_C$. The following elementary fact is used in several papers on this topic [14, 27, 32].

REMARK 3.1. The residual graph G_C resulting at the end of synchronous peeling is the 2-core of G.

It is convenient to reorder the factors (from 1 to m) and variables (from 1 to n) as follows. We index the factors in increasing order according to $F_1, F_2, \ldots, F_{T_C}$, choosing an arbitrary order within each F_t for $1 \leq t \leq T_C$.

For the variable nodes, we first index nodes in V_1, then nodes in V_2 and so on. Within each set V_t, the ordering is chosen in such a way that nodes that have degree 0 in J_{t-1} have lower index than those with degree 1 [notice that, by definition, for any $v \in V_t$, $\deg_{J_{t-1}}(v) \leq 1$]. Finally, for variable nodes in V_t that have degree 1 in J_{t-1}, we use the following ordering. Each such node $v \in V_t$ is connected to a unique factor node in F_t. Call this the *associated factor*, and denote it by f_v. We order the nodes $\deg_{J_{t-1}}(v) = 1$ according to the order of their associated factor, choosing an arbitrary internal order for variable nodes with the same associated factor.

For $A \subseteq F$, $B \subseteq V$, we denote by $\mathbb{H}_{A,B}$ the submatrix of $\mathbb{H}$ consisting of rows with index $a \in A$ and columns $i \in B$. The following structural lemma is immediate, and we omit its proof.

LEMMA 3.2. *Let G be any factor graph [not necessarily in $\mathbb{G}(n, k, m)$] with no 2-core. With the order of factors and variable nodes defined through synchronous peeling, the matrix $\mathbb{H}$ is partitioned in $T_C \times T_C$ blocks $\{\mathbb{H}_{F_s, V_t}\}_{1 \leq s \leq T_C, 1 \leq t \leq T_C}$ with the following structure:*

1. *For any $s > t$, $\mathbb{H}_{F_s, V_t} = 0$.*

2. *The diagonal blocks $\mathbb{H}_{F_s, V_s}$, have a* staircase *structure, namely for each such block the columns can be partitioned into consecutive groups $(\mathcal{C}_l)_{l=0}^{\ell}$, for $\ell = |F_s|$, such that columns in $\mathcal{C}_0$ are equal to 0, columns in $\mathcal{C}_1$ have only the first entry equal to 1, columns in $\mathcal{C}_2$ have only the second entry equal to 1, etc. See below for an example.*

An example of a staircase matrix

(9)
$$\begin{bmatrix} 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Note that V_t is not empty and F_t is not empty for all $t < T_C$. On the other hand, F_{T_C} may be empty, in which case, we adopt the convention that all columns corresponding to V_{T_C} are included in $\mathcal{C}_0$.

The above ordering reduces $\mathbb{H}$ to an essentially upper triangular matrix. It is then immediate to construct a basis for its kernel. We will do this by partitioning the set of variable nodes as the disjoint union $V = U \cup W$ in such a way that

$U \in \{0, 1\}^{m \times m}$ and $\mathbb{H}_U$ is square with full rank, and $W \in \{0, 1\}^{m \times (n-m)}$. We then treat $\underline{x}_W$ as independent variables and $\underline{x}_U$ as dependent ones. The partition is then constructed by letting $W = W_1 \cup \cdots \cup W_{T_C}$ and $U = U_1 \cup \cdots \cup U_{T_C}$, whereby for each $t \in \{1, \ldots, T_C\}$, $W_t \subseteq V_t$ is chosen by considering the staircase structure of block $\mathbb{H}_{F_t, V_t}$ and the corresponding partition over columns $V_t = \mathcal{C}_0 \cup \mathcal{C}_1 \cup \cdots \cup \mathcal{C}_\ell$. We let $W_t = \mathcal{C}_0 \cup \mathcal{C}_1' \cup \cdots \cup \mathcal{C}_\ell'$, where $\mathcal{C}_i'$ includes all the elements of $\mathcal{C}_i$ except the first (and is empty if $|\mathcal{C}_i| = 1$). Finally, $U_t \equiv V_t \setminus W_t$. With these definitions, $\mathbb{H}_{F,U}$ is an $m \times m$ binary matrix with full rank. In addition, it is upper triangular with diagonal blocks $\mathbb{H}_{F_t, U_t} = I_{|U_t|}$ for $t = 1, \ldots, T_C$, where I_r is the $r \times r$ identity matrix.

In order to construct a sparse basis for the clusters when $\alpha > \alpha_d(k)$ [and hence prove Theorem 2, point 2(a)], we will have to consider matrices $\mathbb{H}$ (without a 2-core) that contain rows with exactly 2 nonzero entries (i.e., check nodes of degree 2). Whenever this happens, the construction must be modified, by introducing the notion of *collapsed graph*. The basic idea is that a factor node of degree 2 constrains the adjacent variables to be identical and hence we can replace each set of variables that are thus constrained to be equal by a single proxy variable (a "supernode"). This proxy variable node will have an edge with each factor that was previously connected to a replaced variable node, with a small modification: Since we are operating in $GF(2)$, we retain a single edge for edges with odd multiplicity, and drop edges with even multiplicity.

DEFINITION 3.3.   The *collapsed graph* $G_* = (F_*, V_*, E_*)$ of a graph $G = (F, V, E)$ is the graph of connected components in the subgraph induced by factor nodes of degree 2. Formally,

$$F_* \equiv \{f \in F : |\partial f| \geq 3\},$$
$$V_* \equiv \{S \subseteq V : d_{G^{(2)}}(i, j) < \infty, \forall i, j \in S\},$$
$$E_* \equiv \{(S, a) : S \in V_*, a \in F_*, |\{i \in S \text{ s.t. } (i, a) \in E\}| \text{ is odd}\},$$

where $G^{(2)}$ is the subgraph of G induced by factor nodes of degree 2. We let $n_* \equiv |V_*|$, $m_* \equiv |F_*|$. An element of V_* is referred to as a *supernode*.

Note that for a graph G with no 2-core, the collapsed graph G_* also has no 2-core. We let $\mathbb{Q}$ denote the corresponding adjacency matrix of G_*. Finally, we construct a binary matrix $\mathbb{L}$ with rows indexed by V, and columns indexed by V_*, and such that $L_{i,v} = 1$ if and only if i belongs to connected component v. We apply peeling to G_*, thus obtaining the decomposition of V_* into $U_* \cup W_*$ as described for the original graph G above.

The following is the key deterministic lemma on the construction of the basis. We denote the size of the component of $v \in V_*$ in $G^{(2)}$ by $S(v)$, and for $v \in V_*$, $t \geq 0$ we let $S(v, t) = \sum_{w \in \mathsf{B}_{G_*}(v,t)} S(w)$ be the sum of sizes of vertices within distance t from v.

LEMMA 3.4. *Assume that G_* has no 2-core, then the columns of*

$$\mathbb{L}\begin{bmatrix} (\mathbb{Q}_{F_*,U_*})^{-1}\mathbb{Q}_{F_*,W_*} \\ I_{(n_*-m_*)\times(n_*-m_*)} \end{bmatrix}$$

form an s-sparse basis of the kernel of $\mathbb{H}$, with $s = \max_{v_* \in V_*} S(v_*, T_C(G_*))$. Here, we have ordered the super-nodes $v_* \in V_*$ as U_* followed by W_*, and the matrix inverse is taken over $\mathsf{GF}[2]$.

The proof of Lemma 3.4 is presented in the Appendix A.

3.2. *Construction of the cluster decomposition.* When G does not contain a 2-core [which happens w.h.p. for $\alpha < \alpha_d(k)$] the above lemma is sufficient to characterize the space of solutions $\mathcal{S}$. When G contains a 2-core [w.h.p. for $\alpha > \alpha_d(k)$] we need to construct the partition of the space of solutions $\mathcal{S}_1 \cup \cdots \cup \mathcal{S}_N$.

We let $G_C = (F_C, V_C, E_C)$ denote the 2-core of G, and $P_G : \{0,1\}^V \to \{0,1\}^{V_C}$ be the projector that maps a vector $\underline{x}$ to its restriction $\underline{x}_{V_C}$. Next, we let $\mathbb{H}_C \equiv \mathbb{H}_{F_C,V_C}$ be the restriction of $\mathbb{H}$ to the 2-core, and denote its kernel by $\mathcal{S}_C$. Obviously, for any $\underline{x} \in \mathcal{S}$, we have $P_G\underline{x} \in \mathcal{S}_C$. Further,

(10) $$\mathcal{S} = \bigcup_{\underline{x}_C \in \mathcal{S}_C} \mathcal{S}(\underline{x}_C), \qquad \mathcal{S}(\underline{x}_C) \equiv \{\underline{x} \in \mathcal{S} : P_G\underline{x} = \underline{x}_C\},$$

with $\{\mathcal{S}(\underline{x}_C)\}_{\underline{x}_C \in \mathcal{S}_C}$ forming a partition of $\mathcal{S}$.

It is easy to check $\mathbb{H}_{F\setminus F_C, V\setminus V_C}$ has full row rank. For instance, this follows from the fact that the subgraph induced by $(F \setminus F_C, V \setminus V_C)$ is annihilated by peeling (cf. Remark 3.1). Thus, $\mathcal{S}(\underline{x}_C)$ is nonempty for all $\underline{x}_C \in \mathcal{S}_C$, and the sets $\mathcal{S}(\underline{x}_C)$ are simply translations of each other.

It turns out that $\{\mathcal{S}(\underline{x}_C)\}_{\underline{x}_C \in \mathcal{S}_C}$ is not exactly the partition of $\mathcal{S}$ that we seek. In our next lemma, we show that the set of solutions of the core $\mathcal{S}_C$ can be partitioned in well-separated core-clusters. Moreover, the core-clusters are small and have a high conductance. We will form sets in our partition of $\mathcal{S}$ by taking the union of $\mathcal{S}(\underline{x}_C)$ over $\underline{x}_C$ that lie in a particular core-cluster.

We write $\underline{x}' \preceq \underline{x}$ for binary vectors $\underline{x}', \underline{x}$ if $x'_i \leq x_i$ for all i. We write $\underline{x}' \prec \underline{x}$ if $\underline{x}' \preceq \underline{x}$ and $\underline{x}' \neq \underline{x}$. We need the following definition:

(11) $$\mathcal{L}_C(\ell) \equiv \{\underline{x} : \underline{x} \in \mathcal{S}_C(G), d(\underline{x}, \underline{0}) \leq \ell, \nexists \underline{x}' \in \mathcal{S}_C(G) \setminus \{\underline{0}\} \text{ s.t. } \underline{x}' \prec \underline{x}\}.$$

The set $\mathcal{L}_C(\ell)$ consists of minimal nonzero solutions of the 2-core having weight at most ℓ. (Here, the support of a binary vector $\underline{x}$ is the subset of its coordinates that are nonzero, and the weight of $\underline{x}$ is the size of its support.)

LEMMA 3.5. *For any $\alpha \in (\alpha_d(k), \alpha_s(k))$, there exists $\varepsilon = \varepsilon(\alpha, k) > 0$ such that the following holds. Take any sequence $(s_n)_{n\geq 1}$ such that $\lim_{n\to\infty} s_n = \infty$ and $s_n \leq \varepsilon n$. Let $G \sim \mathbb{G}(n, k, \alpha n)$. Then w.h.p., we have: (i) $\mathcal{L}_C(\varepsilon n) = \mathcal{L}_C(s_n)$; (ii) $|\mathcal{L}_C(\varepsilon n)| < s_n$; (iii) For any $\underline{x}, \underline{x}' \in \mathcal{L}_C(\varepsilon n)$, we have $\underline{x} \wedge \underline{x}' = \underline{0}$, where $\wedge$ denotes bitwise AND. In other words, different elements of $\mathcal{L}_C(\varepsilon n)$ have disjoint supports.*

Lemma 3.5 is proved in Section 7.

REMARK 3.6.   Let E_n be the event that points (i), (ii) and (iii) in Lemma 3.5 hold. Assume E_n and $s_n^2 < \varepsilon n$. Let $\mathcal{S}_{\mathrm{C},1}$ be the set of core solutions with weight less than εn. Then $\mathcal{S}_{\mathrm{C},1}$ forms a linear space over $\mathrm{G}\mathbb{F}(2)$ of dimension $|\mathcal{L}_{\mathrm{C}}(\varepsilon n)|$, with $\mathcal{L}_{\mathrm{C}}(\varepsilon n)$ being a s_n-sparse basis for $\mathcal{S}_{\mathrm{C},1}$. Moreover, every element of $\mathcal{S}_{\mathrm{C},1}$ is s_n^2-sparse.

Let

$$\text{(12)} \qquad\qquad\qquad g \equiv 2^{|\mathcal{L}_{\mathrm{C}}(\varepsilon n)|}.$$

We partition the set $\mathcal{S}_{\mathrm{C}}$ of core solutions in disjoint *core-clusters*, as follows. For $\underline{x}, \underline{x}' \in \mathcal{S}_{\mathrm{C}}$, we write $\underline{x} \simeq \underline{x}'$ if $\underline{x} \oplus \underline{x}' \in \mathrm{span}(\mathcal{L}_{\mathrm{C}}(\varepsilon n))$. It is immediate to see that $\simeq$ is an equivalence relation. We define the core-clusters to be the equivalence classes of $\simeq$. Obviously, the core clusters are affine spaces that differ by a translation, each containing $g \le 2^{s_n}$ solutions. Their number is to be denoted by N. Denote the core-clusters by $\mathcal{S}_{\mathrm{C},1}, \mathcal{S}_{\mathrm{C},2}, \ldots, \mathcal{S}_{\mathrm{C},N}$. Note that for any $\underline{x}, \underline{x}' \in \mathcal{S}_{\mathrm{C}}$ belonging to different core-clusters, we have $d(\underline{x}, \underline{x}') > n\varepsilon$, that is, the core-clusters are well separated. We use the following partition of the solution space (including noncore variables) $\mathcal{S}$ into clusters, based on the core-clusters defined above:

$$\text{(13)} \qquad\qquad \mathcal{S} = \bigcup_{i=1}^{N} \mathcal{S}_i, \qquad \mathcal{S}_i \equiv \{\underline{x} \in \mathcal{S} : P_G \underline{x} \in \mathcal{S}_{\mathrm{C},i}\}.$$

A version of Lemma 3.5 was claimed in [12, 29, 31]. These papers capture the essence of the proof but miss some technical details, and make the erroneous claim that, w.h.p. each pair of core solutions is separated by Hamming distance $\Omega(n)$.

We next want to study the internal structure of clusters. By linearity, it is sufficient to consider only one of them, say $\mathcal{S}_1$, which we can take to contain the origin $\underline{0}$. For any $\underline{x} \in \mathcal{S}_1$, we have $P_G \underline{x} \in \mathcal{S}_{\mathrm{C},1} = \mathrm{span}(\mathcal{L}_{\mathrm{C}}(\varepsilon n))$, and $\mathcal{L}_{\mathrm{C}}(\varepsilon n)$ forms a s_n-sparse basis for $\mathcal{S}_{\mathrm{C},1}$, which coincides with the projection of $\mathcal{S}_1$ onto the core. Consider the subset of solutions $\underline{x} \in \mathcal{S}$, such that $P_G \underline{x} = \underline{x}_{\mathrm{C}}$ for some $\underline{x}_{\mathrm{C}} \in \mathcal{S}_{\mathrm{C},1}$. The set of variables that take the same value for all solutions in this set is strictly larger than the 2-core. In order to capture this remark, we define the *backbone* (variables that are uniquely determined by the core assignment) and *periphery* (other variables) of a graph G.

DEFINITION 3.7.   Define the *backbone augmentation procedure* on G with the initial check induced subgraph $G_{\mathrm{b}}^{(0)}$ as follows. Start with $G_{\mathrm{b}}^{(0)}$. For any $t \ge 0$, pick all check nodes which are not in $G_{\mathrm{b}}^{(t)}$ and have at most one neighbor outside $G_{\mathrm{b}}^{(t)}$. Build $G_{\mathrm{b}}^{(t+1)}$ by adding all these check nodes and their incident edges and neighbors to $G_{\mathrm{b}}^{(t)}$. If no such check nodes exist, terminate and output $G_{\mathrm{b}} = G_{\mathrm{b}}^{(t)}$.

The *backbone* $G_B = (F_B, V_B, E_B)$ of a graph $G = (F, V, E)$ is the output of backbone augmentation procedure on G with the initial subgraph G_C, the 2-core of the graph G.

The *periphery* G_P of a graph $G = (F, V, E)$ is the subgraph induced by the factor nodes $F_P = F \setminus F_B$ and variable nodes $V_P = V \setminus V_B$ that are not in the backbone.[6]

We can now define our basis for S_1. This is formed by two sets of vectors. The first set has a vector corresponding to each element of $\mathcal{L}_C(\varepsilon n)$. For each $\underline{x}_C \in \mathcal{L}_C(\varepsilon n)$, we construct a sparse solution $\underline{x} \in S_1$ such that $P_G \underline{x} = \underline{x}_C$ (Lemma 3.8 below guarantees the existence of such a vector, and bounds its sparsity). This set of vectors forms a basis for the projection of S_1 onto the backbone.

For the second set of vectors, let $\mathbb{H}_P \equiv \mathbb{H}_{F_P, V_P}$ be the matrix corresponding to the periphery graph. We construct a sparse basis for the kernel of the matrix $\mathbb{H}_P$, following the general procedure described in Section 3.1. Namely, we first collapse the graph and then peel it to order the nodes. Note that this second set of basis vectors vanishes on the backbone variables. Lemma 3.4 is used to bound its sparsity.

The first set of vectors is characterized as below (see Section 8 for a proof).

LEMMA 3.8. *Consider any $\alpha \in (\alpha_d(k), \alpha_s(k))$. Let G be drawn uniformly from $\mathbb{G}(n, k, m)$. Take $\varepsilon(\alpha, k) > 0$ from Lemma 3.5, and consider any sequence $(c_n)_{n \geq 1}$ such that $\lim_{n \to \infty} c_n = \infty$. Then, with high probability, the following is true. For every $\underline{x}_C \in \mathcal{L}_C(\varepsilon n)$, there exists c_n-sparse $\underline{x} \in S_1$ such that $P_G \underline{x} = \underline{x}_C$.*

3.3. *Analysis of the construction.* The main challenge in proving Theorem 2 is bounding the sparsity of the bases constructed (either for the full set of solutions, when G does not have a core, or for the cluster S_1, when G has a core). This involves two type of estimates: the first one uses Lemma 3.4, while the second is stated as Lemma 3.8. In the first estimate, we need to bound all the quantities involved in the sparsity upper bound: the number of iterations T after which peeling (on the collapsed graph G_*) halts, and the maximum size $\max_{v \in V_*} S(v, T)$ of any ball of radius T in the collapsed graph. In particular, we will show that, w.h.p., we have $T = O(\log \log n)$, and that $\max_{v \in V_*} S(v, T) \leq (\log n)^C$ w.h.p., which gives sparsity $s \leq (\log n)^C$.

Proving these bounds turns out to be a relatively simpler task when G does not have a 2-core, partly because the graph in question has no factor nodes of degree 2, and thus the collapse procedure is not needed. A second reason is that when G has a 2-core, we need to apply Lemma 3.4 to the periphery subgraph as discussed

---

[6]Notice that there may be a few variables (w.h.p. at most a constant number) in the periphery that also are uniquely determined by the core assignment.

above. Remarkably, the periphery graph admits a relatively explicit probabilistic characterization. We say that a graph is *peelable* if its core is empty, and hence the peeling procedure halts with the empty graph. It turns out that, conditional on the degree distribution, the periphery is uniformly random among all peelable graphs.

Such an explicit characterization is not available, however, when we consider the subgraph obtained by removing the core (the periphery is obtained by removing the entire backbone). Nevertheless, the proof of Lemma 3.8 requires the study of this more complex subgraph. We overcome this problem by using tools from the theory of local weak convergence [5, 6, 9].

Given a graph $G = (F, V, E)$, its check-node degree profile $R = (R_l)_{l \in \mathbb{N}}$ is a probability distribution such that, for any $l \in \mathbb{N}$, $m R_l$ is the number of check nodes of degree l. A degree profile R can conveniently be represented by its generating polynomial $R(x) \equiv \sum_{l \geq 0} R_l x^l$. The derivative of this polynomial is denoted by $R'(x)$. In particular, $R'(1) = \sum_{l \geq 0} l R_l$ is the average degree.

Given integers m, n and a probability distribution $R = (R_l)_{l \leq k}$ over $\{0, 1, \ldots, k\}$, we denote by $\mathbb{D}(n, R, m)$ the set of *check-node-degree-constrained graphs*, that is, the set of bipartite graph with m labeled check nodes, n labeled variable nodes and check node degree profile R. As for the model $\mathbb{G}(n, k, m)$, we will write $G \sim \mathbb{D}(n, R, m)$ to denote a graph drawn uniformly at random from this set. Note that we have restricted the checks to have degree no more than k. Further, we will only be interested in cases with $R_0 = R_1 = 0$.

LEMMA 3.9.   *Let $G = (F, V, E) \sim \mathbb{G}(m, n, k)$ and let G_P be its periphery. Suppose that with positive probability, G_P has n_p variable nodes, m_p check nodes, and check degree profile R^p. Then, conditioned on $G_\mathrm{P} \in \mathbb{D}(n_\mathrm{p}, R^\mathrm{p}, m_\mathrm{p})$, the periphery G_P is distributed uniformly over the set $\mathbb{D}(n_\mathrm{p}, R^\mathrm{p}, m_\mathrm{p}) \cap \mathcal{P}$, where $\mathcal{P}$ is the set of peelable graphs.*

There is a small technical issue here in that if $G' \in \mathbb{D}(n_\mathrm{p}, R^\mathrm{p}, m_\mathrm{p})$, then variable nodes in G' have labels from 1 to n_p, whereas G_P has variable node labels that form a subset of $\{1, 2, \ldots, n\}$, and similarly for check nodes. We adopt the convention that the variable and check nodes in G_P are relabeled sequentially, respecting the original order, before comparing with elements of $\mathbb{D}(n_\mathrm{p}, R^\mathrm{p}, m_\mathrm{p})$.

The above lemma establishes that the periphery is roughly uniform, conditional on being peelable. Its proof is in Section 6.1.

Conceptually, we will bound the sparsity, as estimated in Lemma 3.4 by proceeding in three steps: (1) Bound the estimated basis sparsity $\max_v S(v, T)$ for *check node degree constrained graphs* $\mathbb{D}(n, R, m)$, in terms of the degree distribution; (2) Estimate the "typical" degree distribution for the periphery, and prove concentration around this estimate; (3) Prove that, if R is close to the typical degree distribution, then $G \sim \mathbb{D}(n, R, m)$ is peelable with uniformly positive probability. The latter allows us to transfer the sparsity estimates from the uniform model $\mathbb{D}(n, R, m)$ to the actual distribution of the periphery.

Lemma 3.11 below accomplishes steps (1) and (3), while Lemma 3.12 takes care of step (2). In order to state these lemmas, it is convenient to introduce density evolution (the terminology comes from the analysis of sparse graph codes [27, 28, 36]).

DEFINITION 3.10.   Given $\alpha > 0$, a degree profile R, and an initial condition $z_0 \in [0, 1]$, we define the *density evolution sequence* $\{z_t\}_{t \geq 0}$ by letting for any $t \geq 1$,

$$(14) \qquad\qquad z_t = 1 - \exp\{-\alpha R'(z_{t-1})\}.$$

Whenever not specified, the initial condition will be assumed to be $z_0 = 1$. The one-dimensional recursion (14) will be also called *density evolution recursion*.

We say the pair (α, R) is *peelable at rate η* for $\eta > 0$ if $z_t \leq (1 - \eta)^t / \eta$ for all $t \geq 0$. We say that the pair (α, R) is *exponentially peelable* (for short *peelable*) if there exists $\eta > 0$ such that it is peelable at rate η.

The density evolution recursion (14) describes the large graph asymptotics of a certain belief propagation algorithm that captures the peeling process, and will be described Section 4.

The next lemma is proved in Section 5.

LEMMA 3.11.   *Consider the set $\mathbb{D}(n, R, \alpha n)$, where $R = (R_l)_{l \leq k}$ is a check degree profile such that $R_0 = R_1 = 0$. Assume that the pair (α, R) is peelable at rate η. Then there exist constants $N_0 = N_0(\eta, k) < \infty$, $\delta = \delta(\eta, k) > 0$, $C_1 = C_1(\eta, k) < \infty$, $C_2 = C_2(\eta, k) < \infty$ such that the following hold, for G a random graph drawn from $\mathbb{D}(n, R, m)$ with $n > N_0$:*

(i) *The graph G is peelable with probability at least δ. Further, if $R_2 = 0$, one can take δ arbitrary close to 1 (in other words G is peelable w.h.p.).*
(ii) *Conditional on G being peelable, peeling on the collapsed graph G_* terminates after $T \leq C_1 \log \log n$ iterations, with probability at least $1 - n^{-1/2}$.*
(iii) *Letting $T_{\mathrm{ub}} = \lfloor C_1 \log \log n \rfloor$, we have $\max_{v \in V_*} S(v, T_{\mathrm{ub}}) \leq (\log n)^{C_2}$, with probability at least $1 - n^{-1/2}$.*

Our final lemma is proved in Section 6.2 and establishes the peelability condition for the periphery.

LEMMA 3.12.   *For any $\alpha > \alpha_{\mathrm{d}}$ there exist constants $\eta = \eta(k, \alpha) > 0$, $\gamma_* = \gamma_*(k, \alpha) > 0$ such that the following holds. Let $G = (F, V, E)$ be a graph drawn uniformly at random from the ensemble $\mathbb{G}(n, k, m)$, $m = n\alpha$, and let $G_{\mathrm{P}} = (F_{\mathrm{P}}, V_{\mathrm{P}}, E_{\mathrm{P}})$ be its periphery. Let $m_{\mathrm{P}} \equiv |F_{\mathrm{P}}|$, $n_{\mathrm{P}} \equiv |V_{\mathrm{P}}|$, $\alpha_{\mathrm{P}} \equiv m_{\mathrm{P}}/n_{\mathrm{P}}$ and denote by R^{P} the random check degree profile of G_{P}. Then, for any $\varepsilon > 0$, w.h.p. we have: (i) The pair $(\alpha_{\mathrm{P}}, R^{\mathrm{P}})$ is peelable at rate η; (ii) $n(\gamma_* - \varepsilon) \leq n_{\mathrm{p}} \leq n(\gamma_* + \varepsilon)$.*

3.4. *Putting everything together.*   At this point, we can formally summarize the proof of our main result, Theorem 2, that builds on the construction and analysis provided so far.

PROOF OF THEOREM 2.   1. For $\alpha < \alpha_d(k)$, w.h.p., the graph G does not contain a 2-core (cf. Theorem 4), hence peeling returns an empty graph. Using the construction in Lemma 3.4, we obtain an s-sparse basis, with $s = \max_{v \in V} |\mathsf{B}(v, T_C)|$ (notice that in this case there is no factor node of degree 2, and hence the collapsed graph coincides with the original graph). The number of peeling iterations T_C is bounded by Lemma 3.11(ii), using the fact that, by definition of $\alpha_d(k)$ the pair (α, R), with $R_k = 1$ is peelable at rate $\eta = \eta(\alpha, k) > 0$ for $\alpha < \alpha_d(k)$. Hence, $T_C \leq C_1 \log \log n$ w.h.p., for some $C_1 = C_1(\alpha, k) < \infty$. Finally, by applying Lemma 3.11(iii) we obtain the thesis.

Next, consider point 2. The partition into clusters is constructed as per equation (13), and in particular the number of clusters N is equal to the number of solutions of the core linear system $\mathbb{H}_C \underline{x} = \underline{0}$ divided by g given by equation (12). Let us consider the various claims concerning this partition:

2(a). By construction, it is sufficient to construct a basis of the cluster $\mathcal{S}_1$ containing the origin, cf. Section 3.2. The basis has two sets of vectors.

The first set of vectors is given by Lemma 3.8. Their projection onto the core spans the core solutions in $\mathcal{S}_{C,1}$. Since variables in the backbone are uniquely determined by those on the core, their projection onto the backbone spans the backbone projection of $\mathcal{S}_1$. By Lemma 3.8, these vectors are, w.h.p., c_n-sparse for any $c_n \to \infty$. Lemma 3.4 provides the second set of vectors. These span the kernel of the adjacency matrix of the periphery, $\mathbb{H}_p$ and vanish identically in the backbone. In particular, they are independent from the first set. It is easy to check that the two sets of vectors together form a basis for the cluster $\mathcal{S}_1$.

We are left with the task of proving that the second set of basis vectors is sparse. The construction in Lemma 3.4 proceeds by collapsing the periphery graph G_P, and applying peeling. We thus need to bound the sparsity $s = \max_{v \in V} S(v, T_C)$. Define the event (implicitly indexed by n)

$$\mathsf{E}_1 \equiv \left\{ (\alpha_P, R^P) \text{ is peelable at rate } \eta > 0 \text{ and } n_P \geq n\gamma_*/2 \right\}.$$

By Lemma 3.12, we know that E_1 holds with high probability for suitable choices of $\eta = \eta(k, \alpha) > 0$ and $\gamma_* = \gamma_*(\alpha, k) > 0$. Further $R_0^P = R_1^P = 0$ with probability 1.

From Lemma 3.9, we know that G_P is drawn uniformly from the set $\mathbb{D}(n_P, R^P, m_P) \cap \mathcal{P}$. Let G' be drawn uniformly from $\mathbb{D}(n_P, R^P, m_P)$, with (n_P, R^P, m_P) distributed as for G_P, conditional on $(\alpha_p, R^p) \in \mathsf{E}_1$. We can then apply Lemma 3.11 to G'. From point (i), it follows that G' is peelable with probability at least $\delta = \delta(\alpha, k) > 0$. Let G'_* be the result of collapsing G'. From points (ii) and (iii) it follows that, with probability at least $1 - n_P^{-0.5} \geq 1 - (n\gamma_*/2)^{-0.5} \to 1$ as

$n \to \infty$, we have $\max_{v \in V'_*} S_{G'}(v, T_C) \leq \max_{v \in V'_*} S_{G'}(v, T_{ub}) \leq (\log n)^C$, for some $C = C(\alpha, k) < \infty$. (We use the subscript on S to indicate the graph under consideration.)

Since E_1 holds for G_P w.h.p., and since G' is peelable with probability uniformly bounded away from zero, it follows that the same bound on the sparsity holds for G_P as well. In other words, w.h.p., we have that

$$\max_{v \in V_{P,*}} S_{G_P}(v, T_C) = (\log n)^C.$$

Here, $V_{P,*}$ is the set of super-nodes resulting from the collapse of G_P. Finally, using Lemma 3.4, we deduce that the second set of basis vectors obtained from this construction is s-sparse for $s = (\log n)^C$.

2(b). By Lemma 3.5, w.h.p., for any two core solutions $\underline{x}_C \in \mathcal{S}_{C,1}$, $\underline{x}_C' \in \mathcal{S}_{C,b}$, $b \neq 1$ we have $d(\underline{x}_C, \underline{x}_C') \geq n\varepsilon$. This immediately implies $d(\underline{x}, \underline{x}') \geq n\varepsilon$, for any two solutions $\underline{x} \in \mathcal{S}_1$, $\underline{x}' \in \mathcal{S} \setminus \mathcal{S}_1$. By linearity, we conclude $d(\mathcal{S}_a, \mathcal{S}_b) \geq n\varepsilon$ for all a, b.

2(c). Let N_C be the number of solutions of the core linear system $\mathbb{H}_C \underline{x} = 0$. This was proved to concentrate on the exponential scale in [18, 20], with $n(\Sigma - \varepsilon) \leq \log N_C \leq n(\Sigma + \varepsilon)$ with high probability, and Σ given as in the statement (cf. also [29]). The number of clusters is $N = N_C/g$ for $g = 2^{\mathcal{L}_C(\varepsilon n)}$, cf. equation (12). Using the bound $|\mathcal{L}_C(\varepsilon n)| \leq s_n$ from Lemma 3.5(ii) and choosing s_n to diverge sufficiently slowly with n, we deduce that N also concentrates on the exponential scale with the same exponent as N_C. $\square$

4. A belief propagation algorithm and density evolution.

A useful analysis tool is provided by a belief propagation algorithm [cf. equations (4) and (5)] that refines the peeling algorithm introduced in Section 3.1. The same algorithm is also of interest in iterative coding; see [29, 36].

We restate the BP update rules for the convenience of the reader.

$$v_{v \to a}^t = \begin{cases} *, & \text{if } \widehat{v}_{b \to v}^{t-1} = * \text{ for all } b \in \partial v \setminus a, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$\widehat{v}_{a \to v}^t = \begin{cases} 0, & \text{if } v_{u \to a}^t = 0 \text{ for all } u \in \partial a \setminus v, \\ *, & \text{otherwise.} \end{cases}$$

The initialization at $t = 0$ depends on the context, but it is convenient to single out two special cases. In the first case, all messages are initialized to 0: $v_{v \to a}^0 = \widehat{v}_{a \to v}^0 = 0$ for all $(a, v) \in E$. In the second, they are all initialized to $*$: $v_{v \to a}^0 = \widehat{v}_{a \to v}^0 = *$ for all $(a, v) \in E$. We will refer to these two cases (resp.) as BP_0 and BP_*. We let $\underline{v}^t \equiv (v_{v \to a}^t)_{(a,v) \in E}$ and $\widehat{\underline{v}}^t \equiv (\widehat{v}_{v \to a}^t)_{(a,v) \in E}$ denote the vector of messages.

We mention here that BP_* on the a graph $G \in \mathbb{G}(n, k, m)$ turns out to be trivial (all messages remain $*$). However, we find it useful to run BP_* on the subgraph induced by variable and check nodes outside the core. We describe this in detail in Section 4.2.

The belief propagation algorithm introduced here enjoys an important monotonicity property. More precisely, define a partial ordering between message vectors by letting $0 \succ *$ and $\underline{v} \succeq \underline{v}'$ if $\nu_{v \to a} \succeq \nu'_{v \to a}$ and $\widehat{\nu}_{a \to v} \succeq \widehat{\nu}_{a \to v}$ for all $(a, v) \in E$.

LEMMA 4.1 ([29, 36]). *Given two states $\underline{v}_1^t \succeq \underline{v}_2^t$, we have $\underline{v}_1^{t'} \succeq \underline{v}_2^{t'}$ and $\widehat{\underline{v}}_1^{t'} \succeq \widehat{\underline{v}}_2^{t'}$ at all $t' \geq t$.*

As a consequence, the iteration BP_0 is monotone decreasing (i.e., $\underline{v}^{t+1} \preceq \underline{v}^t$) and BP_* is monotone increasing (i.e., $\underline{v}^{t+1} \succeq \underline{v}^t$). In particular, both converge to a fixed point in at most $|E|$ iterations.

It is not hard to check by induction over t that BP_0 corresponds closely to the peeling process.

LEMMA 4.2. *A variable node v is eliminated in round t of peeling, that is, $v \in V_t$, if there is at most one incoming 0 message to v in iteration $t - 1$ of BP_0 but this was not true in previous rounds. A factor node a is eliminated in round t of peeling (i.e., $a \in F_t$), along with all its incident edges, if it receives a $*$ message for the first time in iteration t of BP_0.*

Further, the fixed point of BP_0 captures the decomposition of G into core, backbone and periphery as follows.

LEMMA 4.3. *Let $(\underline{v}^\infty, \widehat{\underline{v}}^\infty)$ denote the fixed point of BP_0. For $v \in V$, we have*:

- $v \in V_C$ if and only if v receives two or more incoming 0 messages under $\widehat{\underline{v}}^\infty$,
- $v \in V_B \setminus V_C$ if and only if v receives exactly one incoming 0 message under $\widehat{\underline{v}}^\infty$,
- $v \in V_P$ if and only if v receives no incoming 0 messages under $\widehat{\underline{v}}^\infty$.

For $a \in F$, we have

- $a \in F_C$ if and only if a receives no incoming $*$ message under $\underline{v}^\infty$,
- $a \in F_B \setminus F_C$ if and only if a receives one incoming $*$ message under $\underline{v}^\infty$,
- $a \in F_P$ if and only if a receives two or more incoming $*$ messages under $\underline{v}^\infty$.

Finally, G_C is the subgraph induced by (F_C, V_C) and similarly for G_B and G_P.

The proofs of the last two lemmas are based on a straightforward case-by-case analysis, and we omit them. (In fact, this correspondence is well known in iterative coding, albeit in a somewhat different language [36].)

4.1. *Density evolution.* It turns out that distribution of BP messages is closely tracked by density evolution, in the large graph limit. Before stating this fact formally, it is useful to introduce a different ensemble $\mathbb{C}(n, R, m)$ that will be used in some of the proofs. A graph G in $\mathbb{C}(n, R, m)$ is constructed as follows. We label variable nodes 1 through n and check nodes 1 through m. We choose an arbitrary partition of the m check nodes into $k + 1$ sets with the lth set consisting of $m R_l$ check nodes with degree l each, for $l = 0, 1, \ldots, k$. For each check node of degree l, we draw l half-edges distinct from each other. Each of these half-edges is connected to an arbitrary variable node.

There is a close relationship between the sets $\mathbb{D}(n, R, m)$ and $\mathbb{C}(n, R, m)$. Any element of $\mathbb{D}(n, R, m)$ corresponds to $\prod_{l=2}^{k} (l!)^{m R_l}$ elements of $\mathbb{C}(n, R, m)$, with the ambiguity arising due to the ordering of the neighborhood of a check node in $\mathbb{C}(n, R, m)$. Conversely, any element of $\mathbb{C}(n, R, m)$ with no double edges [two or more edges between the same (variable, check) pair] corresponds to a unique element of $\mathbb{D}(n, R, m)$. Moreover, the fraction of elements of $\mathbb{C}(n, R, m)$ that have no double edges is uniformly bounded away from zero as $n \to \infty$ [10]. This leads to Lemma 4.4 below.

LEMMA 4.4. *Let* E *be a graph property that does not depend on edge labels [e.g.,* $\mathsf{E}(G) \equiv \{G$ *is a tree*$\}$*]. There exists* $C = C(k, \alpha_{\max}) < \infty$ *such that the following is true for any* $\alpha \in [0, \alpha_{\max}]$*. Suppose* E *holds with probability* $1 - \varepsilon$ *for G drawn uniformly at random from* $\mathbb{C}(n, R, \alpha n)$*, for some* $\varepsilon \in [0, 1]$*. Then* E *holds with probability at least* $1 - C\varepsilon$ *for* G' *drawn uniformly at random from* $\mathbb{D}(n, R, \alpha n)$.

An important tool in the following will be the notion of almost sure local convergence of graph sequences. We made this notion precise in Definition 1.2, following [15].

We now return to the distribution of BP messages and density evolution.

LEMMA 4.5. *Let* $\{z_t\}$ *be the density evolution sequence defined by* (14)*, for a given polynomial R, with* $z_0 = 1$*, and define* $\widehat{z}_t \equiv R'(z_t)/R'(1)$*. Assume* $G_n \sim \mathbb{D}(n, R, m)$ *or* $G_n \sim \mathbb{C}(n, R, m)$ *with* $m = n\alpha$.

Let $R_{l_0, l_*}^{(t)}$ *be the fraction of check nodes receiving* l_0 *incoming 0 messages and* l_* *incoming* $*$ *messages after t iterations of* BP_0 *in* G_n*. Similarly, let* $L_{l_0, l_*}^{(t)}$ *the fraction of variable nodes receiving* l_0 *incoming 0 messages and* l_* *incoming* $*$ *messages after t iterations of* BP_0.

Then for any fixed $t \geq 0$*, the following occurs almost surely*:

$$(15) \quad \lim_{n \to \infty} R_{l_0, l_*}^{(t)} = R_{l_0 + l_*} \binom{l_0 + l_*}{l_0} z_t^{l_0} (1 - z_t)^{l_*} \quad \text{for } l_0, l_* \in \{0, 1, \ldots, k\},$$

$$(16) \quad \lim_{n \to \infty} L_{l_0, l_*}^{(t)} = \mathbb{P}\{X_0 = l_0, X_* = l_*\} \quad \text{for all } l_0, l_* \in \mathbb{N},$$

where $X_0 \sim \text{Poisson}(R'(1)\alpha\widehat{z}_t)$, $X_* \sim \text{Poisson}(R'(1)\alpha(1 - \widehat{z}_t))$ are two independent Poisson random variables.

PROOF. Notice that both $\mathbb{D}(n, R, m)$ and $\mathbb{C}(n, R, m)$, $m = n\alpha$ converge locally to unimodular bipartite trees. More precisely, if rooted at random variable nodes, they converge to Galton–Watson trees with root offspring distribution $\text{Poisson}(R'(1)\alpha)$ at variable nodes, and equal to the size-biased version of R at check nodes. The proof of the analogous statement in the case of nonbipartite graphs can be found in [15], Proposition 2.6. It uses an explicit calculation to show that the empirical distribution of local neighborhoods converges in expectation, and a martingale concentration argument to verify the assumptions of Borel–Cantelli, and hence deduce almost sure convergence. The same proof extends—with minimal changes—to bipartite (factor) graphs.

Messages are local functions of the graph, hence their distribution converges to the one on the limit tree. In particular, incoming messages on the same node are asymptotically independent because they depend on distinct subtrees. The message distribution can be computed through a standard tree recursion (see [29, 36]) that coincides with the density evolution recursion (14). □

Using the correspondence in Lemma 4.2 between BP_0 and the peeling algorithm, we can use density evolution to track the peeling algorithm.

LEMMA 4.6. *Given a factor graph H, let $n_1(H)$ denote the number of variable nodes of degree 1, and $n_{2+}(H)$ the number of variable nodes of degree 2 or larger in H. For $l \in \mathbb{N}$, let $m_l(H)$ be the number of factor nodes of degree l in H.*

Consider synchronous peeling for $t \geq 1$ rounds on a graph $G \sim \mathbb{D}(n, R, \alpha n)$ or $G \sim \mathbb{C}(n, R, \alpha n)$, with $R_0 = R_1 = 0$, and let J_t denote the residual graph after t iterations. Let $\omega \equiv \alpha R'(1)$. Then for any $\delta > 0$, there exists $N_0 = N_0(\delta, k, t, \alpha)$ such that with probability at least $1 - 1/n^2$

$$\left| \frac{m_l(J_t)}{n} - \alpha R_l z_t^l \right| \leq \delta$$

(17)

$$\text{for } l \in \{2, 3, \ldots, k\},$$

(18) $$\left| \frac{n_1(J_t)}{n} - \omega\widehat{z}_t \exp(-\omega\widehat{z}_t)(1 - \exp(-\omega(\widehat{z}_{t-1} - \widehat{z}_t))) \right| \leq \delta,$$

(19) $$\left| \frac{n_{2+}(J_t)}{n} - 1 + \exp(-\omega\widehat{z}_t)(1 + \omega\widehat{z}_t) \right| \leq \delta.$$

PROOF. For the sake of simplicity, let us consider $n_1(J_t)$. By Lemma 4.2, a node v has degree 1 in the residual graph J_t if and only if there is one incoming 0 message to v at time t, and there were two or more incoming 0 messages to v at time $t - 1$. By Lemma 4.5, the number of incoming 0 messages to v at

time t converges in distribution to $Z_1 \sim \text{Poisson}(\omega \widehat{z}_t)$. Using monotonicity of the algorithm, and again Lemma 4.5, the number of incident edges such that the message incoming to v at time $t - 1$ is 0 but changes to $*$ at time t, converges to $Z_2 \sim \text{Poisson}(\omega(\widehat{z}_{t-1} - \widehat{z}_t))$, and is asymptotically independent of the number of 0 messages (converging to Z_1). Therefore, $n_{1,t}/n$ converges as $n \to \infty$ to

$$\mathbb{P}[Z_1 = 1]\mathbb{P}[Z_2 \geq 1] = \omega \widehat{z}_t \exp(-\omega \widehat{z}_t)\big(1 - \exp(-\omega(\widehat{z}_{t-1} - \widehat{z}_t))\big).$$

This establishes that the estimate (18) holds with high probability. In order to obtain the desired probability bound, one can use a standard concentration of measure argument [19, 36]. Namely, we first condition on the degrees of the check nodes. Since the unconditional distributions $\mathbb{D}(n, R, m)$ and $\mathbb{C}(n, R, m)$ are recovered by a random relabeling of the check nodes, such conditioning is irrelevant. We then regard $n_1(J_t)$ as a function of the independent random variables $X_1, \dots, X_m$ whereby X_a is the neighborhood of the ath check node. We denote by E_n the event that all the balls $\mathsf{B}_G(v, 2t)$ of radius t in G have size smaller than $(\log n)^C$. We have

$$\big|\mathbb{E}\{n_1(J_t)|X_1, \dots, X_{a-1}, X_a; \mathsf{E}_n\} - \mathbb{E}\{n_1(J_t)|X_1, \dots, X_{a-1}, X'_a; \mathsf{E}_n\}\big| \leq (\log n)^C.$$

The desired probability estimate then follows by applying Azuma's inequality (in a form that allow for exceptional events; see, e.g., [19], Theorem 7.7) and bounding $\mathbb{P}(\mathsf{E}_n^c)$ (see, e.g., Section 5.2). $\square$

4.2. *BP fixed points.* For our purposes, it is important to characterize the fixed point of the BP_0 algorithm introduced above. Indeed, the structure of this fixed point is directly related to the decomposition of G into core, backbone and periphery (cf. Lemma 4.3), which is in turn crucial for our definition of clusters. Let us start from an easy remark on density evolution.

LEMMA 4.7. *Let $\{z_t\}_{t \geq 0}$ be the density evolution sequence defined by equation (14) with initial condition $z_0 = 1$. Then $t \mapsto z_t$ is monotone decreasing, and hence has a limit $Q \equiv \lim_{t \to \infty} z_t$ which is given by*

$$(20) \qquad\qquad Q = \sup\{z \text{ s.t. } z = 1 - \exp\{-\alpha R'(z)\}\}.$$

PROOF. Monotonicity follows from the fact that $z \mapsto f(z) \equiv 1 - \exp\{-\alpha R'(z)\}$ is monotone increasing, and that $z_1 = 1 - \exp\{-\alpha R'(1)\} < z_0$, whence $z_2 = f(z_1) \leq f(z_0) = z_1$, and so on. $\square$

Notice that the definition of Q given in this lemma is consistent with the one in Theorem 1, that corresponds to the special case of regular, degree-k check nodes, that is, $R(x) = x^k$. We further let $\widehat{Q} \equiv R'(Q)/R'(1)$.

We know that both BP_0 and density evolution converge to a fixed point. Since density evolution tracks BP_0 for any bounded number of iterations, it would be

tempting to conclude that a description of the BP_0 fixed point is obtained by replacing z_t by Q and $\widehat{z}_t$ by $\widehat{Q}$ in Lemma 4.5. This is, of course, far from obvious because it requires an inversion of the limits $n \to \infty$ and $t \to \infty$. Despite this caveat, this substitution is essentially correct.

LEMMA 4.8. *Assume $G_n \sim \mathbb{G}(n, k, m)$ with $m = n\alpha$, and $\alpha \in [0, \alpha_{\mathrm{d}}(k)) \cup (\alpha_{\mathrm{d}}(k), \infty)$.*

Let $R_{l_0, l_*}^{(\infty)}$ be the fraction of check nodes receiving l_0 incoming 0 messages and l_* incoming $*$ messages at the fixed point of BP_0. Similarly, let $L_{l_0, l_*}^{(\infty)}$ the fraction of variable nodes receiving l_0 incoming 0 messages and l_* incoming $*$ messages at the fixed point of BP_0.

The following occurs with probability 1:

$$(21) \quad \lim_{n \to \infty} R_{l_0, l_*}^{(\infty)} = \binom{k}{l_0} Q^{l_0} (1 - Q)^{l_*} \quad \text{for } l_0 \in \{0, 1, \dots, k\}, l_* = k - l_0,$$

$$(22) \quad \lim_{n \to \infty} L_{l_0, l_*}^{(\infty)} = \mathbb{P}\{X_0 = l_0, X_* = l_*\} \quad \text{for all } l_0, l_* \in \mathbb{N},$$

where $X_0 \sim \mathrm{Poisson}(k\alpha\widehat{Q})$, $X_* \sim \mathrm{Poisson}(k\alpha(1 - \widehat{Q}))$ are two independent Poisson random variables.

Given Lemma 4.5 above, Lemma 4.8 says that the messages change very little beyond a large constant number of iterations. A hint at the fact that Lemma 4.8 is significantly more challenging than Lemma 4.5 is given by the assumption in the former that $\alpha \neq \alpha_{\mathrm{d}}(k)$. In fact, this turns out to be a necessary assumption, because it implies an important correlation decay property.

Molloy [32] established the analog of equation (22) for $\sum_{\ell_0 \geq 2, \ell_* \geq 0} L_{l_0, l_*}^{(\infty)}$, which corresponds to the relative size of the core. We find that the complete theorem presents new challenges: keeping track of the backbone turns out to be hard. One hurdle is that the "estimated backbone" after t iterations of BP_0 (i.e., the subset of variable nodes that receive exactly one 0 message) does not evolve monotonically in t. In contrast, the "estimated core" (i.e., the subset of variable nodes that receive two or more 0 messages) can only shrink. Another hurdle is that, unlike the periphery (cf. Section 6), it turns out that the backbone is *not* uniformly random conditioned on the degree sequence.

The proof of Lemma 4.8 is quite long and will be presented in Section 4.3. The basic idea is to run BP starting from the initialization with 0 messages coming from vertices in the core and $*$ messages everywhere else. This corresponds to BP_* on the noncore G_{NC} [i.e., the subgraph induced by $(F \setminus F_{\mathrm{C}}, V \setminus V_{\mathrm{C}})$], since messages outside the noncore do not change: Messages within the core and from core variables to noncore checks stay fixed to 0. Messages from noncore checks to core variables stay fixed to *. We refer to this algorithm simply as BP_*, with the understanding that BP_* is actually run on G_{NC}.

It is not hard to check by induction over t that BP_* corresponds to the backbone augmentation procedure.

LEMMA 4.9. *Consider the backbone augmentation procedure with the initial subgraph G_C. A factor node a is added to the backbone in round t of backbone augmentation, that is, $a \in G_b^{(t)} \setminus G_b^{(t-1)}$ (cf. Definition 3.7) if all but one incoming message to a in iteration t of BP_* are 0, but this was not the case in previous iterations.*

A variable node v is added to the backbone in round t, of backbone augmentation, that is, $v \in G_b^{(t)} \setminus G_b^{(t-1)}$ if there is one incoming 0 message to v in iteration t of BP_* but this was not true in previous iterations.

It then follows immediately from Lemma 4.3 that BP_0 and BP_* converge to the same fixed point. Denote the messages at this fixed point by $\nu_{v \to a}^{0, \infty}$.

Denote by $\nu_{v \to a}^{*, t}$ the messages produced in iteration t of BP_*, and $\nu_{v \to a}^{0, t}$ the messages produced by BP_0. Monotonicity of BP update implies $\nu_{v \to a}^{0, t} \succeq \nu_{v \to a}^{0, \infty} \succeq \nu_{v \to a}^{*, t}$. The proof consists in showing that the fraction of 0 messages in $\{\nu_{v \to a}^{0, t}\}_{(a, v) \in E}$ is, for large fixed t, close to the fraction of 0 messages in $\{\nu_{v \to a}^{*, t}\}_{(a, v) \in E}$. The challenge is that no analog of Lemma 4.5 is available for BP_*.

Our final lemma is a straightforward consequence of Lemmas 4.5 and 4.8 above.

LEMMA 4.10. *Consider any $k \geq 3$, any $\alpha \in (0, \alpha_d) \cup (\alpha_d, \alpha_s)$ and any $\delta > 0$. There exists $T < \infty$ such that the following occurs. Let $G_n \sim \mathbb{G}(n, k, \alpha n)$. Then, eventually (in n) almost surely, the fraction of (check-to-variable or variable-to-check) messages that change after iteration T of BP_0 is smaller than δ.*

PROOF. Let $N^t(n)$ be the fraction of variable-to-check messages that are equal to 0 after t iterations on G_n (with $t = \infty$ corresponding to the fixed point). Then equations (15) and (21) imply that

$$\left| N^t(n) - z_t \right| \leq \frac{\delta}{3k}, \qquad \left| N^\infty(n) - Q \right| \leq \frac{\delta}{3k}$$

holds eventually almost surely. Using Lemma 4.7, there exists T large enough so that, for $t \geq T$, $|z^t - Q| \leq \delta/(3k)$. By the triangle inequality $|N^t(n) - N^\infty(n)| \leq \delta/k$. The thesis for variable-to-check messages follows since, by monotonicity of BP_0, $N^t(n) - N^\infty(n)$ is exactly equal to the fraction of messages that change value from iteration t to the fixed point. Each change in a variable-to-check message can lead to a change in at most $k - 1$ check-to-variable messages. Thus, the fraction of check-to-variable messages that change after iteration T is smaller than δ. $\square$

4.3. *Proof of Lemma* 4.8. Throughout this section, the notion of convergence adopted is *convergence locally* (cf. Definition 1.2).

For $n \geq 0$, draw a graph G_n uniformly at random from $\mathbb{G}(n, k, \alpha n)$. Consider equation (22). Since the total number of incoming messages is equal to the vertex degree, which is Poisson($k\alpha$), it is sufficient to control the distribution of 0 incoming messages. In particular, we define

$$L_{\ell+}^{(t)} \equiv \sum_{\ell_*=0}^{\infty} \sum_{l_0=\ell}^{\infty} L_{l_0, l_*}^{(t)},$$

that is the fraction of nodes that receive ℓ or more 0 incoming messages.

We prove a series of lemmas, leading to the desired estimate for $L_{\ell+}^{(t)}$.

An upper bound on $L_{\ell+}^{(\infty)}$ is relatively easy to obtain.

LEMMA 4.11. *With probability* 1 *with respect to the choice of* $(G_n)_{n \geq 0}$, *we have for all* $l \geq 0$,

$$\lim_{n \to \infty} \sup L_{\ell+}^{(\infty)} \leq \mathbb{P}\{\text{Poisson}(k\alpha \widehat{Q}) \geq \ell\}.$$

PROOF. Using Lemma 4.5 (and using the fact that $L_l \leq C \exp(-l/C)$ for all l holds eventually almost surely, for some $C < \infty$) we have,

$$\lim_{n \to \infty} L_{\ell+}^{(t)} = \mathbb{P}\{\text{Poisson}(k\alpha \widehat{z}_t) \geq \ell\}$$

holds w.p. 1. From Lemma 4.1, it follows that $L_{\ell+}^{(t)}$ is monotone decreasing. Thus, we have

$$\lim_{n \to \infty} \sup L_{\ell+}^{(\infty)} = \mathbb{P}\{\text{Poisson}(k\alpha \widehat{z}_t) \geq \ell\}$$

w.p. 1.

Fix an arbitrary $\delta > 0$. Lemma 4.7 implies that, for t large enough,

$$[\mathbb{P}\{\text{Poisson}(k\alpha \widehat{z}_t) \geq \ell\} - \mathbb{P}\{\text{Poisson}(k\alpha \widehat{Q}) \geq \ell\}] \leq \delta,$$

which implies that

$$\lim_{n \to \infty} \sup L_{\ell+}^{(\infty)} \leq \mathbb{P}\{\text{Poisson}(k\alpha \widehat{Q}) \geq \ell\} + \delta$$

holds almost surely. Since δ is arbitrary, we obtain the claimed result. □

The lower bound on $L_{\ell+}^{(\infty)}$ cannot be obtained by the same approach. We go therefore through a detour.

Let $\mu_n \equiv \mu(G_n)$ be the measure on rooted factor graphs with marks (called "networks" in [5]), constructed as follows: Choose a uniformly random variable node $i \in V_n$ as root. Mark variable nodes with mark c if they are in the 2-core of G_n.

LEMMA 4.12. *The sequence $\{\mu_n\}_{n \geq 0}$ converges locally to the measure on random rooted tree with marks, $\mathcal{T}_*(\alpha, k)$, defined as follows. Construct a random bipartite Galton–Watson tree rooted at $\varnothing$ with offspring distribution* Poisson$(k\alpha)$ *at variable nodes and deterministic $k - 1$ at factor notes. Let $V_C(\mathcal{T}_*)$ be the maximal subset of its vertices such that each variable node has degree at least 2 and each factor node has degree k in the induced subgraph. Mark with* c *all vertices in $V_C(\mathcal{T}_*)$.*

PROOF.   It is immediate to see that the sequence $\{\mu_n\}_{n \geq 0}$ is tight almost surely with respect to the choice of $(G_n)_{n \geq 0}$, that is, that for any $\varepsilon \geq 0$ there exists a compact set $\mathcal{K}$ such that $\mathbb{P}\{H_*(n) \in \mathcal{K}\} \geq 1 - \varepsilon$. (E.g., take $\mathcal{K}$ to be the set of graphs that have maximum degree Δ_t at distance t for a suitable sequence $t \mapsto \Delta_t$.) Therefore [5], any subsequence of $\{\mu_n\}$ admits a further subsequence that converges locally weakly to a limiting measure on rooted networks. This subsequence can be constructed through a diagonal argument: First, construct a subsequence $\{\mu_{n_s^t}\}_{s \geq 0}$ such that the depth-t subtree converges. Refine it to get a subsequence $\{\mu_{n_s^{t+1}}\}_{s \geq 0}$ such that the depth-$(t + 1)$ subtree converges and so on. Finally, extract the diagonal subsequence $\{\mu_{n_s^s}\}_{s \geq 0}$.

We will prove the thesis by a standard weak convergence argument [24]: We will show that for any subsequence of $\{\mu_n\}_{n \geq 0}$, there is a sub-subsequence that converges locally weakly to the measure on $\mathcal{T}_*(\alpha, k)$.

Consider indeed any sub-subsequence that converges locally weakly to limiting random rooted graph with marks, which we denote by $\mathcal{O}_*$. Define the *unmarking* operator U that maps a marked rooted graph to the corresponding unmarked rooted graph. We have that $\mathsf{U}(\mathcal{O}_*) \stackrel{\mathrm{d}}{=} \mathsf{U}(\mathcal{T}_*)$ (here $\stackrel{\mathrm{d}}{=}$ denotes equality in distribution) from local weak convergence of random graphs to Galton–Watson trees (see, e.g., [6, 15]). We will hereafter couple the two trees in such a way that $\mathsf{U}(\mathcal{O}_*) = \mathsf{U}(\mathcal{T}_*)$.

Recall that a stopping set is any subset of variable nodes of a factor graph, such that each variable node has degree at least 2 in the induced subgraph. The 2-core of the factor graph is the maximal stopping set and is a superset of any stopping set. These notions are well defined for infinite graphs as well.

Now, the marks in $\mathcal{T}_*$ correspond to the core by definition. The marks in $\mathcal{O}_*$ form a stopping set, since the measure on $\mathcal{O}_*$ is the local weak limit of μ_n, and in any graph drawn from μ_n, w.p. 1 a vertex is marked only if at least two of its neighboring checks have all marked neighboring variable nodes. Moreover, one can show that both $\mathcal{T}_*$ and $\mathcal{O}_*$ are unimodular. Indeed $\mathcal{T}_*$ is unimodular since the unmarked tree is clearly unimodular, and the marking process does not make any reference to the root. Unimodularity of $\mathcal{O}_*$ is clear since it is the local weak limit of a marked random graph [5]. Thus, in order to prove our thesis it suffices to show that the density of marks is the same in $\mathcal{T}_*$ and $\mathcal{O}_*$. (Because the subset of nodes that is marked in $\mathcal{T}_*$ contains the subset marked in $\mathcal{O}_*$ and the density of

their difference is equal to the difference of the densities. Finally, for unimodular network, if a mark type has density 0, then the set of marked nodes is empty by union bounds.)

Let

$$\mathsf{E} \equiv \left\{ \lim_{n \to \infty} |V_{\mathrm{c}}(G_n)|/n = \mathbb{P}\{\mathrm{Poisson}(k\alpha \widehat{Q}) \geq 2\} \right\},$$

where Q and $\widehat{Q}$ are defined as at the beginning of Section 4. It was proved in [32] that $|V_{\mathrm{c}}(G_n)|/n \xrightarrow{\text{a.s.}} \mathbb{P}\{\mathrm{Poisson}(k\alpha \widehat{Q}) \geq 2\}$, that is, the event E occurs with probability 1. Now let the set of marked vertices in $\mathcal{O}_*$ be denoted by $\widehat{V}_{\mathrm{C}}(\mathcal{O}_*)$. It is easy to see that if E holds, the density of marks in $\mathcal{O}_*$ is given by

$$(23) \qquad\qquad \mathbb{P}\{\varnothing \in \widehat{V}_{\mathrm{C}}(\mathcal{O}_*)\} = \mathbb{P}\{\mathrm{Poisson}(k\alpha \widehat{Q}) \geq 2\}.$$

Proceeding analogously to the proof of [8], Proposition 1.2, we obtain

$$(24) \qquad\qquad \mathbb{P}\{\varnothing \in V_{\mathrm{C}}(\mathcal{T}_*)\} = \mathbb{P}\{\mathrm{Poisson}(k\alpha \widehat{Q}) \geq 2\}.$$

The sketch of this step is the following. Let E_t be the event that $\varnothing$ belongs to a "depth t core," where the requirement of "degree at least 2 in the subgraph" applies only to variables up to depth $t - 1$. The probability on the left-hand side is just $\mathbb{P}\{\mathsf{E}\}$ for $\mathsf{E} = \bigcap_{t \geq 1} \mathsf{E}_t$. Since E_t is a decreasing sequence, $\mathbb{P}\{\mathsf{E}\} = \lim_{t \to \infty} \mathbb{P}\{\mathsf{E}_t\}$. On the other hand, $\mathbb{P}\{\mathsf{E}_t\}$ can be computed explicitly through a tree calculation and converges to $\mathbb{P}\{\mathrm{Poisson}(\alpha k \widehat{Q}) \geq 2\}$ as $t \to \infty$ yielding (24).

Finally, the thesis follows by comparing equations (23) and (24), and recalling that $\mathbb{P}(\mathsf{E}) = 1$. □

We next construct a random tree $\widetilde{\mathcal{T}}_*(\alpha, k)$ with marks on the directed edges as follows. Marks take values in $\{0, *\}$ and to each undirected edge we associate a mark for each of the two directions. We will refer to the direction toward the root as to the "upward" direction, and to the opposite one as to the "downward" direction. The marks correspond to fixed point BP messages, and we will call them messages as well in what follows. First, consider only edges directed upward. This is a multitype GW tree. At the root generate $\mathrm{Poisson}(k\alpha)$ offsprings, and mark each of the edges to 0 independently with probability $\widehat{Q}$, and to $*$ otherwise. At a nonroot variable node, if the parent edge is marked 0, generate $\mathrm{Poisson}(k\alpha(1 - \widehat{Q}))$ descendant edges marked $*$ and $\mathrm{Poisson}_{\geq 1}(k\alpha \widehat{Q})$ descendant edges marked 0 [here $\mathrm{Poisson}_{\mathsf{E}}(\lambda)$ denotes a Poisson random variable with parameter λ conditional to E]. If the parent edge is marked $*$, generate $\mathrm{Poisson}(k\alpha(1 - \widehat{Q}))$ descendant edges marked $*$ and no descendant edges marked 0. At a factor node, if the parent edge is marked 0, generate $k - 1$ descendant edges marked 0. If the parent node is marked $*$, generate $M \sim \mathrm{Binom}_{\leq k-2}(k - 1, Q)$ descendants marked 0, and $k - 1 - M$ descendants marked $*$.

For edges directed downward, marks are generated recursively following the usual BP rules, cf. equations (4), (5), starting from the top to the bottom. It is easy

to check that with this construction, the marks in $\widetilde{\mathcal{T}}_*(\alpha, k)$ correspond to a BP fixed point.

We extend the unmarking operator U by allowing it to act on graphs with marks on edges (and removing the marks).

LEMMA 4.13. $\mathsf{U}(\widetilde{\mathcal{T}}_*)$ and $\mathsf{U}(\mathcal{T}_*)$ have the same distribution.

PROOF. For this, we construct $\mathsf{U}(\widetilde{\mathcal{T}}_*)$ (which is $\widetilde{\mathcal{T}}_*$ without the marks revealed) in a "breadth first" manner as follows: First, we draw a Poisson(αk) number of factor descendants for the root node. Let a be a factor descendant of the root. Then a has $k - 1$ variable node descendants. The message $\widehat{v}_{a \to \varnothing}$ is 0 with probability $\widehat{Q}$. It immediate to check from our construction and $\widehat{Q} = Q^{k-1}$ that:

Fact 1: Conditional on the degree of the root $\deg(\varnothing) = d_1$, the $d_1(k - 1)$ upward messages incoming to the check nodes $a \in \partial\emptyset$ are independent, with $\mathbb{P}\{v_{v \to a} = 0\} = Q$.

Now, we draw the number of descendants for each neighbor of a. Using fact 1, together with the definition of $\widetilde{\mathcal{T}}$, one can check that:

Fact 2: Conditional on the degree of the root $\deg(\varnothing) = d_1$, the number of descendants of each of the $d_1(k - 1)$ variable nodes v at the first generation is an independent Poisson($k\alpha$) random variable. Further, the upward messages toward these variable nodes are independent with $\mathbb{P}\{\widehat{v}_{b \to v} = 0\} = \widehat{Q}$.

This argument (outlined for simplicity for the first generation) can be repeated almost verbatim at any generation. Denote by $\widetilde{\mathcal{T}}_{*,d}$ the first d generations of $\widetilde{\mathcal{T}}_{*,d}$ (with variable nodes at the leaves). One then proves by induction that at any d, conditional on $\mathsf{U}(\widetilde{\mathcal{T}}_{*,d})$, the number of descendants of the variable nodes in the last generation are i.i.d. Poisson($k\alpha$), and given these, the corresponding upward messages are i.i.d. $\mathbb{P}\{\widehat{v}_{b \to v} = 0\} = \widehat{Q}$. This implies the thesis. □

LEMMA 4.14. $\widetilde{\mathcal{T}}_*$ is unimodular.

PROOF. We already established unimodularity of $\mathsf{U}(\widetilde{\mathcal{T}}_*)$ [since $\mathsf{U}(\widetilde{\mathcal{T}}_*) = \mathsf{U}(\mathcal{T}_*)$ is a unimodular Galton–Watson tree]. To establish the claim, let $\widetilde{\mathcal{T}}'_*$ be the random tree whose distribution has Radon–Nikodym derivative $\deg(\varnothing)/\mathbb{E}\{\deg(\varnothing)\}$ with respect to that of $\widetilde{\mathcal{T}}_*$. We need to show that moving the root to a uniformly random descendant variable node of the root (via one check) in $\widetilde{\mathcal{T}}'_*$, leaves the distribution of $\widetilde{\mathcal{T}}'_*$ unchanged (cf. [5], Section 4).

Draw $\widetilde{\mathcal{T}}'_*$ at random, weighted by the degree of the root $\varnothing$. In this argument, we make the root explicit by denoting the tree by $(\widetilde{\mathcal{T}}'_*, \varnothing)$. Reveal the degree $d_1 = \deg(\varnothing)$ of the root. We have $d_1 > 0$ almost surely. Take a uniformly random neighboring check $a \in \partial\varnothing$, and a uniformly random descendant i of a (we know that a has $k - 1$ descendants). Reveal the number of descendants of i. Let this number be $d_2 - 1$, so that i has d_2 neighbors in total. Note that we do not reveal any of the messages in $\widetilde{\mathcal{T}}'_*$. At this point, consider the incoming messages

to the variable nodes $\varnothing$ and i except for $\widehat{v}_{a \to \varnothing}$ and $\widehat{v}_{a \to i}$, and the incoming messages to the check a except for $v_{\varnothing \to a}$ and $v_{i \to a}$. Call this vector of messages M. The messages in M are independent, with probability $\widehat{Q}$ of for each incoming message to variable nodes to be 0, and probability Q for incoming messages to a to be 0.[7] The messages $\widehat{v}_{a \to \varnothing}$, $\widehat{v}_{a \to i}$, $v_{\varnothing \to a}$ and $v_{i \to a}$ are deterministic functions of M. Finally, notice that d_1 and d_2 are independent, and identically distributed as $1 + \mathrm{Poisson}(\alpha k)$. At this point, it is clear that $(\widetilde{\mathcal{T}}_*', i)$ is distributed identically to $(\widetilde{\mathcal{T}}_*', \varnothing)$, which establishes unimodularity. $\quad\square$

LEMMA 4.15.   *Let* F *be a map from "trees with marked edges" to "trees with marked variable nodes" defined as follows*: $\mathsf{F}(\mathcal{T})$ *is obtained from* $\mathcal{T}$ *by putting a* c *mark on vertex i if and only if at least two incoming edges have a 0 mark.*
    Then $\mathsf{F}(\widetilde{\mathcal{T}}_*(\alpha, k)) \stackrel{\mathrm{d}}{=} \mathcal{T}_*(\alpha, k)$.

PROOF.   It is easy to check that the subset of variable nodes in $\widetilde{\mathcal{T}}_*$ that receive two or more incoming 0's forms a stopping set (since the set of messages is at a BP fixed point). But the density of marked nodes in $\mathsf{F}(\widetilde{\mathcal{T}}_*)$ (i.e., the probability of the root being marked) is $\mathbb{P}\{\mathrm{Poisson}(k\alpha\widehat{Q}) \geq 2\}$, which is exactly the same as the density of marked nodes in $\mathcal{T}_*$ (recall that $\mathcal{T}_*$ is also unimodular, cf. proof of Lemma 4.12). On the other hand, the set of marked nodes in $\mathcal{T}_*$ is the core by definition and hence includes the marked nodes in $\mathsf{F}(\widetilde{\mathcal{T}}_*)$. We deduce that the set of vertices that are marked in $\mathcal{T}_*$ but not in $\mathsf{F}(\widetilde{\mathcal{T}}_*)$ has vanishing density and, therefore, $\mathsf{F}(\widetilde{\mathcal{T}}_*(\alpha, k)) \stackrel{\mathrm{d}}{=} \mathcal{T}(\alpha, k)$. $\quad\square$

We let B be the subset of variable nodes v of $\widetilde{\mathcal{T}}_*(\alpha, k)$ such that at least one message incoming to v is equal to 0. Then this set has density

$$(25) \qquad \mathbb{P}\{\varnothing \in B\} = \mathbb{P}\{\mathrm{Poisson}(k\alpha\widehat{Q}) \geq 1\} \equiv \widehat{Q}.$$

In light of Lemma 4.15, we further denote the set of variable nodes in $\widetilde{\mathcal{T}}_*$ having two or more incoming 0 messages by $V_{\mathrm{C}}(\widetilde{\mathcal{T}}_*)$.

Consider running BP_* on $\mathsf{U}(\widetilde{\mathcal{T}}_*)$ [this is BP starting with zeros from the variable nodes in $V_{\mathrm{C}}(\widetilde{\mathcal{T}}_*)$ and $*$ elsewhere]. Let the trees with marks on edges obtained after t iterations be denoted by $\widetilde{\mathcal{T}}_*^t$.

Denote by $\tilde{\mu}_n^t$ the measure on the rooted factor graph with marks on the edges constructed as follows: Choose a uniformly random variable node $i \in V(G_n)$. Mark the edges (in each direction) with the messages corresponding to BP_* run for t iterations.

LEMMA 4.16.   *The measures* $(\tilde{\mu}_n^t)_{n \geq 0}$ *converge locally to the measure on* $\widetilde{\mathcal{T}}_*^t$.

---

[7]The argument establishing this is essentially the one above, where we showed that $\mathsf{U}(\widetilde{\mathcal{T}}_*) = \mathsf{U}(\mathcal{T}_*)$.

PROOF. This result is immediate from Lemmas 4.12 and 4.15. □

The following is immediate from the construction of $\widetilde{\mathcal{T}}_*$.

REMARK 4.17. If $\varnothing \in B$, then there exists a subtree of $\widetilde{\mathcal{T}}_*$ rooted at $\varnothing$ with the following properties: (i) If j is a variable node in the subtree, either $j \in V_{\mathrm{C}}(\widetilde{\mathcal{T}}_*)$ or at least one descendant factor node is in the subtree; (ii) If a is a factor node in the subtree, all its descendants are also in the subtree.

We call the subtree just defined a *witness* for $\varnothing$ (there might be more than one in principle). Notice that a priori a witness can be finite [if it ends up with nodes in $V_{\mathrm{C}}(\widetilde{\mathcal{T}}_*)$], or infinite.

LEMMA 4.18. *Almost surely any node $i \in B$ has a finite witness. Thus, $\lim_{t \to \infty} \widetilde{\mathcal{T}}_*^t = \widetilde{\mathcal{T}}_*$.*

PROOF. It is sufficient to prove that the following event has zero probability: $\varnothing \in B$ and $\varnothing$ only has infinite witnesses. Suppose $\varnothing \in B$. We will look for a minimal witness for $\varnothing$. If $\varnothing \in V_{\mathrm{C}}(\widetilde{\mathcal{T}}_*)$, then it is itself a witness and we are done. If not then, there is exactly one incoming 0 message, say from factor a. Then factor a has $k - 1$ incoming 0 messages from descendants. The subtrees corresponding to these descendants are independent. Consider a descendant i of a. We have

$$\mathbb{P}\big(i \in B \setminus V_{\mathrm{C}}(\widetilde{\mathcal{T}}_*)\big) = \mathbb{P}\{\mathrm{Poisson}_{\geq 1}(\alpha k \widehat{Q}) = 1\}$$
$$= \exp(-\alpha k \widehat{Q}) \alpha k \widehat{Q} / (1 - \exp(-\alpha k \widehat{Q}))$$
$$= \exp(-\alpha k \widehat{Q}) \alpha k Q^{k-2}.$$

Conditioned on $i \in B \setminus V_{\mathrm{C}}(\widetilde{\mathcal{T}}_*)$, the node i has exactly $k - 1$ descendant variable nodes (via one check node). Thus, conditioned on $\varnothing \in B$, the minimal witness is a Galton–Watson tree with offspring distributed as Z, whereby $Z = (k - 1)$ with probability $\exp(-\alpha k \widehat{Q}) \alpha k Q^{k-2}$, and $Z = 0$ otherwise. The branching factor of this tree is $\exp(-\alpha k \widehat{Q}) \alpha k (k - 1) Q^{k-2} < 1$ (cf. Lemma 6.6 below). The lemma follows. □

LEMMA 4.19. *Consider the setting of Lemma 4.8. We have*

$$\lim_{n \to \infty} \inf L_{1+}^{(\infty)} \geq \mathbb{P}\{\mathrm{Poisson}(k \alpha \widehat{Q}) \geq 1\},$$

almost surely with respect to the choice of G_n.

PROOF. Let B_t be the subset of variable nodes in $\widetilde{\mathcal{T}}_*^t$ that receive at least one 0 message. Let y_t be the density of nodes in B_t. From Lemma 4.18, we have immediately

(26) $$\lim_{t \to \infty} y_t = \mathbb{P}\{\mathrm{Poisson}(k \alpha \widehat{Q}) \geq 1\}.$$

Let $B_t(n) \subseteq V(G_n)$ be the subset of nodes having at least one incoming 0 after t iterations of BP_*. Let $y_t(n)$ be the fraction of these nodes, that is, $y_t(n) \equiv |B_t(n)|/n$. From Lemma 4.16, we have

$$(27) \qquad \lim_{n \to \infty} y_t(n) = y_t$$

almost surely. By equation (26), we have $\lim_{n \to \infty} y_t(n) \geq \mathbb{P}\{\mathrm{Poisson}(k\alpha\widehat{Q}) \geq 1\} - \delta$ for all $t \geq T(\delta)$. By monotonicity of BP_*, we have $\liminf_{n \to \infty} L_{1+}^{(\infty)} \geq \lim_{n \to \infty} y_t(n) \geq \mathbb{P}\{\mathrm{Poisson}(k\alpha\widehat{Q}) \geq 1\} - \delta$, which implies the thesis. $\square$

LEMMA 4.20. *Consider the setting of Lemma* 4.8. *We have, for all* $\ell \geq 2$,

$$\liminf_{n \to \infty} L_{\ell+}^{(\infty)} \geq \mathbb{P}\{\mathrm{Poisson}(k\alpha\widehat{Q}) \geq \ell\},$$

almost surely with respect to the choice of G_n.

PROOF. The proof is very similar to that of the previous lemma. Let $C(\ell; n) \subseteq V$ be the subset of variable nodes in G_n that are in the core and have at least ℓ neighboring check nodes in the core. Then we have (by monotonicity of BP_*)

$$(28) \qquad L_{\ell+}^{(\infty)} \geq \frac{|C(\ell; n)|}{n}.$$

On the other hand, let $y(\ell)$ be the density of variable nodes in $\widetilde{\mathcal{T}}_*$ that receive two or more 0 messages and have at least ℓ neighboring check nodes in the set

$$\{a : \text{For each } i \in \partial a, \text{ node } i \text{ receives two or more 0 messages}\}.$$

It follows from Lemmas 4.12 and 4.15 that

$$(29) \qquad \liminf_{n \to \infty} \frac{1}{n}|C(\ell; n)| = y(\ell).$$

On the other hand, it is easy to check that the construction of $\widetilde{\mathcal{T}}_*$ implies that $y(\ell)$ coincides with the density of nodes receiving ℓ or more 0 messages (here the assumption $\ell \geq 2$ is crucial). Hence, $y(\ell) = \mathbb{P}\{\mathrm{Poisson}(k\alpha\widehat{Q}) \geq \ell\}$, which together with equations (28), (29) yields the thesis. $\square$

PROOF OF LEMMA 4.8. Equation (22) follows from Lemmas 4.11, 4.19 and 4.20. Equation (21) follows from a completely analogous argument. $\square$

Recall that $\tilde{\mu}_n^t$ is the measure on the rooted factor graph with marks on the edges constructed as follows: Choose a uniformly random variable node $i \in V(G_n)$. Mark the edges (in each direction) with the messages corresponding to BP_* run for t iterations. Recall that $\tilde{\mu}_n^*$ is defined similarly with marks corresponding to the BP fixed point. Denote by $\tilde{\mu}_n^t(d)$, the measure obtained from $\tilde{\mu}_n^t$ by restricting the depth of the rooted graph to d.

LEMMA 4.21. *For any $d \geq 0$ and any $\delta > 0$, there exists $t < \infty$ such that almost surely,*

$$\limsup_{n \to \infty} \|\tilde{\mu}_n^t(d) - \tilde{\mu}_n^*(d)\|_{\mathrm{TV}} < \delta.$$

PROOF. Consider running BP_* on G_n. From Lemma 4.16, we know $\tilde{\mu}_n^t$ converges locally to the measure on $\widetilde{\mathcal{T}}_*^t$. From Lemma 4.18, we know $\lim_{t \to \infty} \widetilde{\mathcal{T}}_*^{(t)} = \widetilde{\mathcal{T}}_*^\infty$. In particular, the fraction of 0 variable-to-check messages in $\widetilde{\mathcal{T}}_*^{(t)}$ converges to Q (i.e., the fraction of 0 variable-to-check messages in $\widetilde{\mathcal{T}}_*^{(\infty)}$). But from Lemma 4.8, the fraction of 0 variable-to-check messages in $\tilde{\mu}_n^*$ converges eventually almost surely to the same value, and similarly for check-to-variable messages the fraction of 0 messages converges to $\widehat{Q}$ [using the fact that $L_l \leq C \exp(-l/C)$ for all l holds eventually almost surely, for some $C < \infty$]. Using monotonicity of BP_*, we deduce that for any $\varepsilon > 0$, there exists t large enough such that,

(30) $\limsup_{n \to \infty} \{\text{Number of message changes after iteration } t \text{ in } G_n\}/n \leq \varepsilon$

holds almost surely. Now, we can choose ε small enough such that eventually (in n) almost surely, for any set of εn edges in G_n, the union of balls of radius d around these edges contains no more than δn nodes. Combining with equation (30), at least $(1 - \delta)$ fraction of nodes have all messages in a ball of radius d unchanged after iteration t, almost surely. This yields the result. $\square$

PROOF OF THEOREM 3. From Lemma 4.16, we know $\tilde{\mu}_n^t$ converges locally to the measure on $\widetilde{\mathcal{T}}_*^t$. From Lemma 4.18, we know $\lim_{t \to \infty} \widetilde{\mathcal{T}}_*^t = \widetilde{\mathcal{T}}_*^\infty$. Combining with Lemma 4.21, we obtain that

(31) $$\limsup_{n \to \infty} \|\tilde{\mu}_n^*(d) - \mu(\widetilde{\mathcal{T}}_*^\infty(d))\|_{\mathrm{TV}} < \delta$$

almost surely. Since δ is arbitrary, we obtain, for every d, that

(32) $$\limsup_{n \to \infty} \|\tilde{\mu}_n^*(d) - \mu(\widetilde{\mathcal{T}}_*^\infty(d))\|_{\mathrm{TV}} = 0$$

holds almost surely. The result follows. $\square$

5. Proof of Lemma 3.11: Peelability implies a sparse basis.

5.1. *Proof of Lemma 3.11(i) and (ii).* Let us begin by describing the proof strategy.

Instead of analyzing peeling on the collapsed graph G_*, we analyze a different peeling process. We first run synchronous peeling on G for a large constant τ number of iterations. We then collapse the resulting graph, as discussed in Section 3.1, that is, coalescing variables connected to each other via degree 2 factors (cf. Definition 3.3). Finally, we run synchronous peeling on the collapsed graph

until it gets annihilated. We show that this process takes at least as many iterations as synchronous peeling on G_* (Lemma 5.1 below). In order to bound the number of iterations under this new two-stages process, we proceed as follows. We choose the constant τ such that the residual graph J_τ is subcritical, and hence consists of trees and unicyclic components of size $O(\log n)$ w.h.p. As a consequence, the collapsed graph—to be denoted by $\mathsf{T}(J_\tau)$—contains only checks of degree 3 or more, and consists of trees and unicyclic components of size $O(\log n)$. It is not hard to show that it takes only $O(\log\log n)$ additional rounds of peeling to annihilate $\mathsf{T}(J_\tau)$ under this condition (see Lemma 5.4 below).

Several technical lemmas follow, which are proved in the Appendix B, except Lemma 5.1, which we prove below. At the end of the subsection, we provide a proof of Lemma 3.11, parts (i) and (ii).

Consider the peeling algorithm and define J to be the peeling operator corresponding to one round of synchronous peeling (cf. Table 1). Thus, for a bipartite graph G, the residual graph after t rounds of peeling is $J^t(G)$. Denote by $J^\infty(G)$ the graph produced by the peeling procedure after it halts: this is the empty graph if G is peelable, and the core of G otherwise. Recall that $T_{\mathsf{C}}(G)$ denotes the number of rounds of peeling performed before halting at $J^\infty(G)$. Further, define T to be the collapse operator as per Definition 3.3. For instance $G_* = \mathsf{T}(G)$. The next lemma bounds from above the number of rounds of peeling required to annihilate G_*, in terms of the modified peeling process (consisting of τ rounds of peeling, followed by collapse, and then peeling until annihilation).

LEMMA 5.1. *For any constant $\tau \geq 0$ and any peelable bipartite graph G,*

$$T_{\mathsf{C}}\big(\mathsf{T}(G)\big) \leq T_{\mathsf{C}}\big(\mathsf{T}\big(J^\tau(G)\big)\big) + \tau.$$

Peelability of a pair (α, R) immediately implies some useful properties.

LEMMA 5.2. *For a factor degree profile (α, R) that is peelable at rate $\eta > 0$, we have*:

(i) $2\alpha R_2 \leq 1 - \eta$.
(ii) $\alpha \leq 1$.

Notice that the factor graph induced by degree 2 check nodes is in natural correspondence with an ordinary graph (replace every check node by an edge) which is uniformly random given the number of edges. The average degree of this graph is $2\alpha R_2$, and Lemma 5.2(i) implies that it is subcritical, as we would expect for a peelable degree distribution.

Lemma 3.11 is stated for the ensemble $\mathbb{D}(n, R, m)$, $m = n\alpha$. However, in parts of the proof of this lemma, we find it convenient to work instead with the ensemble $\mathbb{C}(n, R, m)$ introduced in Section 4.1.

We need to characterize the residual graph J_t after t rounds of peeling. Lemmas 5.3 and 4.6 achieves this for $G \sim \mathbb{C}(n, R, m)$. Together, they show essentially that density evolution provides an accurate characterization of J_t. Using these Lemmas, we are able to deduce [see proof of Lemma 3.11(i) and (ii) below] that J_τ consists of small trees and unicyclic components w.h.p., for large enough τ. Finally, using Lemma 4.4, we apply the same results to $G \sim \mathbb{D}(n, R, m)$.

Recall that $n_1(G)$ denotes the number of variable nodes of degree 1 in G, and $n_{2+}(G)$ denotes the number of variable nodes of degree 2 or more in G. Let

$$(33) \qquad \mathbb{C}(n, R, m; n_1', n_2') \equiv \{G : G \in \mathbb{C}(n, R, m), n_1(G) = n_1', n_{2+}(G) = n_2'\}.$$

In the lemma below, we slightly modify the peeling process, choosing to retain all variable nodes V in the residual graph (check nodes are eliminated as usual). With a slight abuse of notation, we keep denoting by J_t the residual graph, although this is obtained from J_t by adding a certain number of isolated variable nodes.

LEMMA 5.3.   *Consider a graph G drawn uniformly at random from $\mathbb{C}(n, R, m)$. For any $t \in \mathbb{N}$, consider synchronous peeling for t rounds on G, resulting in the residual graph J_t. Suppose that for some $(\widetilde{R}, \tilde{m}, \tilde{n}_1, \tilde{n}_2)$, we have $J_t \in \mathbb{C}(n, \widetilde{R}, \tilde{m}; \tilde{n}_1, \tilde{n}_2)$ with positive probability. Then, conditioned on $J_t \in \mathbb{C}(n, \widetilde{R}, \tilde{m}; \tilde{n}_1, \tilde{n}_2)$, the residual graph J_t is uniformly random within $\mathbb{C}(n, \widetilde{R}, \tilde{m}; \tilde{n}_1, \tilde{n}_2)$.*

Our final technical lemma bounds the number of peeling rounds needed to annihilate a tree or unicyclic component.

LEMMA 5.4.   *Consider a factor graph $G = (F, V, E)$ with no check nodes of degree 1 or 2, and that is a tree or unicyclic. Then G is peelable and $T_C(G) \leq 2\lceil \log_2 |V| \rceil$.*

PROOF OF LEMMA 3.11(i) and (ii).   A standard calculation (see, e.g., [14], or Section 7.1 which carries through a similar calculation) shows that, for a uniformly random graph $\mathbb{C}(n, \widetilde{R}, \tilde{m}; \tilde{n}_1, \tilde{n}_2)$, with $\tilde{n}_1, \tilde{n}_2 \geq n\varepsilon$ and with $\tilde{m}\widetilde{R}'(1) \geq \tilde{n}_1 + 2\tilde{n}_2 + n\varepsilon$ for some $\varepsilon > 0$, the asymptotic degree distribution of variable nodes is

$$\mathbb{P}\{D = 0\} = q_0,$$

$$\mathbb{P}\{D = 1\} = q_1,$$

$$\mathbb{P}\{D = \ell\} = (1 - q_0 - q_1)\mathbb{P}\{\text{Poisson}_{\geq 2}(\lambda) = \ell\} \qquad \text{for all } \ell \geq 2$$

for suitable choices of q_0, q_1, λ depending on the ensemble parameters. Further, by a standard breadth-first search argument, the neighborhood of a vertex v is dominated stochastically by a (bipartite) Galton–Watson tree, with offspring distribution

equal to the size-biased version of $\widetilde{R}$ at check nodes, and equal to of $\mathbb{P}\{D = \cdot\}$ at variable nodes.

Consider $G \sim \mathbb{C}(n, R, m)$. Using Lemmas 4.6 and 5.3, it is possible to estimate the degree distribution, of J_t. A lengthy but straightforward calculation shows that the corresponding branching factor is $\theta(J_t) = \alpha R'(z_t)$. Now, notice that

$$R'(z) = 2R_2 + \sum_{l=3}^{k} l(l-1)z^{l-2} \leq 2R_2 + k(k-1)z$$

for $z \leq 1$. Choose $\tau = \tau(\eta, k) < \infty$ such that $z_\tau \leq \eta/(3\alpha k(k-1))$. Then we have $\alpha R'(1)\rho'(z_\tau) \leq 2\alpha R_2 + \eta/3$. But Lemma 5.2 tells us that $2\alpha R_2 \leq 1 - \eta$. It follows that $\alpha R'(z_\tau) \leq 1 - 2\eta/3$.

In particular, the branching factor $\theta = \theta(J_\tau)$ associated with the random graph J_τ satisfies $\theta \leq 1 - \eta/3$, with probability at least $1 - 1/n^2$. Following a standard argument [11] where we explore the neighborhood of v by breadth first search, we obtain that with probability at least $1 - 1/n^{1.7}$ for $n \geq N_1(\eta, k)$, the connected component containing v is a tree or unicyclic, with size less than $C_4 \log n$, for some $C_4 = C_4(\eta, k) < \infty$. Applying a union bound, we obtain that for $n \geq N_2 = N_2(\eta, k)$, with probability at least $1/n^{0.7}$, the event E_n occurs, where

(34)
$$\mathsf{E}_n \equiv \{\text{All connected components in } J_\tau \text{ are trees or unicyclic}$$
$$\text{and have size at most } C_4 \log n\}.$$

Then, from Lemma 4.4, we infer that E_n occurs with probability at least $1/n^{0.6}$ for $G \sim \mathbb{D}(n, R, m)$ provided $n \geq N_3$, where $N_3 = N_3(k) < \infty$. We stick to $G \sim \mathbb{D}(n, R, m)$ for the rest of this proof.

We now analyze the peeling process starting with J_τ and consider only what happens on E_n since it occurs with sufficiently large probability. Let us consider first point (i). Clearly, tree components are peelable. If $R_2 = 0$, then there are no factors of degree 2, and unicyclic components are also peelable (Lemma 5.4). Thus, the entire graph is annihilated by peeling w.h.p., as claimed. If $R_2 > 0$, then the number of unicyclic components of size smaller than M is asymptotically Poisson with parameter $C_5 < \infty$ uniformly bounded in M (this follows, e.g., by [37]; see also [11, 38]). It follows that with probability at least $\exp(-C_5)/2$ for $n \geq N_4$, there are no unicyclic components of size smaller than M. The expected number of unicyclic components of size M or larger is upper bounded by $\sum_{\ell \geq M} \theta^\ell/(2\ell) \leq \theta^M/(1-\theta)$, and for M large enough no unicyclic component of this sizes exists, with probability at least $1 - \exp(-C_5)/4$. Considering these two contributions, the graph contains no cycle with probability at least $\exp(-C_5)/4$ for $n \geq N_4$, and hence it is peelable. This completes part (i).

For (ii), notice that in collapsing a connected component of J_τ, the number of variable nodes does not increase. Further, a tree component collapses to a tree and a unicyclic component collapses either to a tree or a unicyclic components. Thus, we can use Lemma 5.4 with $N \leq C_4 \log n$ to obtain the a bound of

$(C_1/2) \log\log n \leq C_1 \log\log n - \tau$ on the number of additional peeling rounds needed, with probability at least $1 - 1/n^{0.6}$. Since the probability of peelability is uniformly bounded away from zero as $n \to \infty$, the probability that the same bound on the number of peeling rounds holds conditioned on peelability is at least (for some $\delta > 0$) $1 - 1/(\delta n^{0.6}) \geq 1 - 1/n^{0.5}$ for $n \geq N_5$, as required. $\square$

5.2. *Proof of Lemma* 3.11(iii).   The following lemma bounds the size of a supercritical Galton–Watson tree, observed up to finite depth. The proof is in the Appendix B.

LEMMA 5.5.   *Consider a Galton–Watson branching process* $\{Z_t\}_{t=0}^{\infty}$ *with* $Z_0 = 1$ *and with offspring distribution* $\mathbb{P}\{Z_1 = j\} = b_j$, $j \geq 0$. *Suppose* $b_r \leq (1 - \delta)^r/\delta$ *for all* $r \geq 0$, *for some* $\delta > 0$. *Also, assume that the branching factor satisfies* $\theta \equiv \sum_{j=1}^{\infty} j b_j = \mathbb{E}[Z_1] > 1$. *Then there exists* $C = C(\delta) > 0$ *such that the following happens.*

For any $\beta > 3$ *and* $T \in \mathbb{N}$, *we have*

$$(35) \qquad \mathbb{P}\left[\sum_{t=0}^{T} Z_t > (\beta\theta)^T\right] \leq 2\exp\left(-C(\beta/3)^T\right).$$

PROOF OF LEMMA 3.11(iii).   From Lemma 5.2(ii), we know that $\alpha \leq 1$. The following occurs in the collapse process: Let $G^{(2)} = (F^{(2)}, V, E^{(2)})$ be the subgraph of G induced by the degree 2 factor nodes (with isolated vertices retained). We have $F_* = F \setminus F^{(2)}$. All variable nodes that belong to a single connected component of $G^{(2)}$ coalesce into a single supernode $v' \in V_*$ in G_*, with a neighborhood that consists of the union of the individual neighborhoods restricted to F_* (cf. Definition 3.3). As mentioned above, $G^{(2)}$ is a random factor graph with $\alpha R_2 n$ factor nodes of degree 2, and is in one-to-one correspondence with a uniformly random graph. For $v' \in V_*$, we denote by $S(v')$ the number of variable nodes in V in the component v'. Lemma 5.2(i) implies that the branching factor of $G^{(2)}$ obeys $2\alpha R_2 \leq 1 - \eta$, that is, $G^{(2)}$ is subcritical. This leads to the following claim that follows immediately from a well-known result on the size of the largest connected component in a subcritical random graph [11].

Claim 1: There exists $C_2 = C_2(\eta) < \infty$, $N_2 = N_2(\eta) < \infty$ such that the following occurs for all $n > N_2$. No component $v' \in V_*$ is composed of more than $C_2 \log n$ variable nodes, that is, $\max_{v' \in V_*} S(v') \leq C_2 \log n$, with probability at least $1 - 1/n$.

Let $G^{\sim 2} \equiv (F_*, V, E \setminus E^{(2)})$, that is, $G^{\sim 2}$ is the subgraph of G induced by factors of degree greater than 2 (with isolated vertices retained).

From Poisson estimates on the node degree distribution, we get the following.

Claim 2: There exists $C_3 = C_3(\eta, k) < \infty$, $N_3 = N_3(\eta, k) < \infty$ such that the following occurs. For all $n > N_3$, no variable node $v \in V$ has degree larger than

$C_3 \log n$ in $G^{\sim 2}$, that is, $\deg_{G^{\sim 2}}(v) \le C_3 \log n$ for all $v \in V$, with probability at least $1 - 1/n$.

Note that we used $\alpha < 1$ [from Lemma 5.2(i)] to avoid dependence on α in the above claim.

Let

$$\mathsf{E}_n \equiv \{ S(v') \le C_2 \log n \text{ for all } v' \in V_* \}$$
$$\cap \{ \deg_{G^{\sim 2}}(v) \le C_3 \log n \text{ for all } v \in V \}.$$

Using claims 1 and 2 above and a union bound, we deduce that E_n holds with probability at least $1 - 2/n$ for $n > N_4$, for some $N_4 = N_4(\eta, k) < \infty$.

Clearly, $G^{\sim 2}$ is independent of $G^{(2)}$. In particular, for $v \in V$ that is part of supernode $v' \in V_*$, we know that $|S(v')|$ is independent of $G^{\sim 2}$. There is a slight dependence between the degree of different variable nodes, but assuming E_n, the effect of this is small if we only condition on $\mathrm{polylog}(n)$ nodes in G_*. This enables our bound on the size of balls in G_*.

Recall that the distribution of random variable X_1 is dominated by the distribution of X_2, if there exists a coupling between X_1 and X_2 such that $X_1 \le X_2$ with probability 1. In bounding the size of a ball of radius T_{ub}, we are justified in replacing degree distributions by dominating distributions, and in assuming that there are no loops.

Fixing a vertex $v \in V_*$, we construct the ball $\mathsf{B}_{G_*}(v, T_{\mathrm{ub}})$ sequentially through a breadth-first search. Choose $\varepsilon = \eta/2$. For n large enough, the distribution of $|S(v')|$ is dominated by the distribution of the number of nodes in a Galton–Watson tree with offspring distribution $\mathrm{Poisson}(2\alpha R_2 + \varepsilon)$. The distribution of $\deg_{G^{\sim 2}}(v)$ is dominated by $\mathrm{Poisson}(\alpha(\sum_{l=3}^{k} l R_l) + \varepsilon)$. In particular, the degree distribution of G_* is dominated by a geometric distribution $b_r \le (1 - \delta)^r/\delta$ for some $\delta = \delta(\eta, k) > 0$. Assuming E_n, this also holds conditionally on the nodes revealed so far, as long as the number of these is, say, $\mathrm{polylog}(n)$.

Thus, assuming E_n, the number of nodes in a ball of radius $T_{\mathrm{ub}} = C_1 \log \log n$ is dominated by the number of nodes in a Galton–Watson tree of depth T_{ub} with offspring distribution $(b_r)_0^\infty$ satisfying $b_r \le (1 - \delta)^r/\delta$ for some and $\theta \equiv \sum_{j=1}^\infty j b_j < C_5$, for $n \ge N_5$. We deduce from Lemma 5.5 that

$$(36) \qquad \mathbb{P}\Big[\max_{v' \in V_*} |\mathsf{B}_{G_*}(v', T_{\mathrm{ub}})| \le (\log n)^{C_6} \Big| \mathsf{E}_n\Big] \ge 1 - 1/n$$

for some $C_6 = C_6(\eta, k) < \infty$, where $|\mathsf{B}_{G_*}(v', T_{\mathrm{ub}})|$ denotes the number of supernodes in $\mathsf{B}_{G_*}(v', T_{\mathrm{ub}})$. But given E_n, the size of components $v' \in V_*$ is uniformly bounded by $C_2 \log n$. Thus, conditioned on E_n, we have $\max_{v' \in V_*} S(v', T_{\mathrm{ub}}) \le C_2(\log n)^{C_6+1}$ with probability at least $1 - 1/n$. At this point, we recall that $\mathbb{P}[\mathsf{E}_n] > 1 - 2/n$, and the result follows. $\square$

6. Characterizing the periphery. Consider a factor graph G when it has a nontrivial 2-core. Recall the definitions of the 2-core, backbone and periphery of a graph from Section 3.2. First, we note some of the properties of these subgraphs that will be useful in the proof of the main lemmas of this section.

As a matter of notation, for a bipartite graph G chosen uniformly at random from the set $\mathbb{G}(n, k, m)$ we denote by G_P the periphery of G and by G_p (lower case subscript) a subgraph of G that is a potential candidate for being the periphery of G. Similarly, we denote by G_B the backbone of G and by G_b a subgraph of G that is a potential candidate for being the backbone of G.

6.1. *Proof of Lemma 3.9: Periphery is conditionally a uniform random graph.* Lemma 3.9 states that if we fix the number of nodes and the check degree profile of the periphery of a graph G chosen uniformly at random from the set $\mathbb{G}(n, k, m)$ then the periphery, G_P, is distributed uniformly at random conditioned on being peelable. Since the original graph G is chosen uniformly at random, in order to prove this lemma it is enough to count, for each possible choice of the periphery G_P, the number of graphs G that have the periphery G_P.

Before proving Lemma 3.9, we first introduce the concept of a "rigid" graph and establish a monotonicity property for the backbone augmentation procedure which was defined in Section 3.2. We use the notation $G \subseteq G'$ if G is a subgraph of G'.

LEMMA 6.1. *Let $G = (F, V, E)$ be a bipartite graph and let G_s be the subgraph of G induced by some $F_s \subseteq F$. Let F_l and F_u be subsets of F such that $F_l \subseteq F_u$ and $F_l \subseteq F_s$. Let $B_l^{(0)}$ be the subgraph induced by F_l (so $B_l^{(0)} \subseteq G_s$) and let $B_u^{(0)}$ be the subgraph induced by F_u. Denote by $B_l^{(\infty)}$ the output of the backbone augmentation process on G_s with the initial graph $B_l^{(0)}$ and by $B_u^{(\infty)}$ the output of the backbone augmentation process on G with the initial graph $B_s^{(0)}$. Then $B_l^{(\infty)} \subseteq B_u^{(\infty)}$.*

The proof of Lemma 6.1 can be found in the Appendix C.

DEFINITION 6.2. Define a graph to be *rigid* if its backbone is the whole graph. We denote by $\mathcal{R}(n, k, m)$ the class of rigid graphs with n variable nodes, and m check nodes each of degree k.

LEMMA 6.3. *Consider a bipartite graph $G = (F, V, E)$ from the ensemble $\mathbb{G}(n, k, m)$. For some set of check nodes $F_b \subseteq F$ denote by $G_b = (F_b, V_b, E_b)$ the subgraph induced by F_b, and denote by $G_p = (F_p, V_p, E_p)$ the subgraph of G induced by the pair $(F_p \equiv F \setminus F_b, V_p \equiv V \setminus V_b)$. Assume G_b and G_p satisfy the following conditions*:

- G_p *is peelable,*
- G_b *is rigid,*
- $|\partial a| \geq 2, \forall a \in F_p$.

Then G_b is the backbone of G (and G_p is the periphery).

PROOF. If G_b is empty, the lemma is trivially true. Assume G_b is nonempty. We prove this lemma in two steps. In the first step we prove that G_b is a subgraph of G_B, the backbone of G. In the second step we show that G_B cannot contain anything outside G_b.

Since G_b is rigid, it contains a nonempty 2-core $(G_b)_C$ and the output of the backbone augmentation procedure with initial graph $(G_b)_C$ is G_b itself. Furthermore, $(G_b)_C$ is part of G_C, the 2-core of the original graph G, since by definition a 2-core is the maximal stopping set (cf. Definition 2.2) and $(G_b)_C$ is a stopping set in G. Hence, the monotonicity of the backbone augmentation procedure implies that $G_b \subseteq G_B$.

In the second step, we prove that G_B cannot contain any node outside G_b. First, note that G_p cannot contain any check node from the 2-core of the original graph G. We prove this by contradiction. Suppose instead that $\tilde{F}$ is the nonempty set of all the check nodes from the 2-core of G that are in G_p. Let $\tilde{V}$ be the set of neighbors of $\tilde{F}$ in G_p. The nodes in $\tilde{V}$ are also part of the 2-core of G and have degree at least 2 in the 2-core of G. Furthermore, there is no edge incident from F_b to V_p because, by definition, G_b is check-induced. In particular, in the 2-core of G, there is no other edge incident on variables in $\tilde{V}$ beyond the ones coming from $\tilde{F}$. Hence, in the nonempty subgraph $\tilde{G} \subseteq G_p$ induced by the check nodes in $\tilde{F}$ and all their neighbors every variable node has degree at least 2. This subgraph is then, by definition, a stopping set in G_p. But by assumption G_p is peelable and cannot contain a stopping set. This is a contradiction that rules out the existence of a nonempty set $\tilde{F}$. Hence, the 2-core of G is contained entirely in G_b (recall that both G_b and the 2-core are check-induced).

Let $B^{(G_C)}$ and $B^{(G_b)}$ be the output of the backbone augmentation procedure on G, once with initial subgraph given by the 2-core of G and once with the initial subgraph given by G_b (which contains the 2-core of G). By monotonicity, $B^{(G_C)} \subseteq B^{(G_b)}$. But the process with the initial subgraph G_b terminates immediately since, by assumption, all check node outside G_b have at least two neighbors in G_p. Therefore, $G_B = B^{(G_C)} \subseteq B^{(G_b)} = G_b$. This completes our proof. □

It is easy to see that the converse of Lemma 6.3 is also true, as stated below.

REMARK 6.4. If $G_b = G_B$ is the backbone of G, then the subgraphs G_b and $G_p = G \setminus G_b = G_P$ satisfy the condition of Lemma 6.3. Here, $G \setminus G_b$ denotes the subgraph of G induced by $(F \setminus F_b, V \setminus V_b)$.

Notice that the fact that the graph $G \setminus G_b$ is peelable follows from the connection between the peeling algorithm and BP_0 stated in Lemmas 4.2 and 4.3. We stated that the messages coming out of the backbone are always 0. From the check node update rule, an incoming 0 message to a check node can be dropped without changing any of the outgoing messages as long as there is at least one other incoming message. By definition, there is no edge between variable nodes in the periphery and check nodes in the backbone. Furthermore, all the check nodes in the periphery have at least two neighbors in the periphery. Therefore, BP_0 on the periphery has the same messages as the corresponding messages of BP_0 on the whole graph. In particular, the fixed point of BP_0 on the periphery is all $*$ messages which shows that the periphery subgraph is peelable. We now prove Lemma 3.9.

PROOF OF LEMMA 3.9.   Our goal is to characterize the probability of observing the periphery of G to be $G_p = (F_p, V_p, E_p)$. We use the shorthand notation $G \setminus G_p$ to denote the subgraph of G induced by the check-variable nodes pair $(F \setminus F_p, V \setminus V_p)$. Let $G_b = (F \setminus F_p, V \setminus V_p, E_b) = G \setminus G_p$ and $E_{pb} = \{(i, a) | i \in V \setminus V_p, a \in F_p\}$ be a set of edges that satisfy the condition $\deg_{E_p}(a) + \deg_{E_{pb}}(a) = k$ for all $a \in F_p$. As before, we denote by G_B and G_P the actual periphery and backbone of the graph G. Define the set of rigid graphs on n_b variable nodes, m_b check nodes and check degree k, $\mathcal{R}(n_b, k, m_b)$, as

(37)
$$\begin{aligned} &\mathcal{R}(n_b, k, m_b) \\ &= \{G_b = (F_b, V_b, E_b) : |F_b| = m_b, V_b = n_b, |\partial a| = k \; \forall a \in F_b, G_b \text{ is rigid}\}. \end{aligned}$$

By Lemma 6.3,

(38)
$$\begin{aligned} &\{G \in \mathbb{G}(n, k, m) : G_P = G_p, G_B = G_b\} \\ &= \{G \in \mathbb{G}(n, k, m) : G_p \subseteq G, G \setminus G_p = G_b, G_p \in \mathcal{P}, G_b \in \mathcal{R}\}, \end{aligned}$$

and in particular,

(39)
$$\begin{aligned} &\{G \in \mathbb{G}(n, k, m) : G_P = G_p\} \\ &= \{G \in \mathbb{G}(n, k, m) : G_p \subseteq G, G_p \in \mathcal{P}, G \setminus G_p \in \mathcal{R}\}. \end{aligned}$$

From equation (39), and counting all the choices for the subgraph $G_b = G \setminus G_p$, and the edges that connect G_p and G_b,

(40)
$$\begin{aligned} &|\{G \in \mathbb{G}(n, k, m) : G_P = G_p\}| \\ &= \sum_{G_b} \sum_{E_{pb}} |\{G \in \mathbb{G}(n, k, m) : G_p \subseteq G, G_p \in \mathcal{P}, G \setminus G_p = G_b, G_b \in \mathcal{R}, \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad E \setminus (E_p \cup E_b) = E_{pb}\}|. \end{aligned}$$

IBRAHIMI, KANORIA, KRANING AND MONTANARI

For fixed G_p and G_b, we can count the number of ways these two subgraphs can be connected to each other. Letting $\bar{R}$ be the degree profile of G_p, we have

(41)
$$\begin{aligned}
&\big|\{G \in \mathbb{G}(n, k, m) : G_P = G_p\}\big| \\
&= \sum_{G \setminus G_p} \prod_{l=2}^{k} \binom{n - |V_p|}{k - l}^{|F_p|\bar{R}_l} \mathbb{I}(G \setminus G_p \in \mathcal{R}) \mathbb{I}(G_p \in \mathcal{P}).
\end{aligned}$$

We can rewrite this as

(42)
$$\begin{aligned}
&\big|\{G \in \mathbb{G}(n, k, m) : G_P = G_p\}\big| \\
&= \prod_{l=2}^{k} \binom{n - |V_p|}{k - l}^{|F_p|\bar{R}_l} \big|\mathcal{R}(n - |V_p|, k, m - |F_p|)\big| \mathbb{I}(G_p \in \mathcal{P}).
\end{aligned}$$

It is clear that the cardinality of the set $\mathcal{R}(n_b, k, m_b)$ is a function of only n_b and m_b. Hence,

(43) $$\big|\{G \in \mathbb{G}(n, k, m) : G_P = G_p\}\big| = Z(n_p, k, R^p, m_p) \mathbb{I}(G_p \in \mathcal{P}),$$

for some function $Z(\cdot, \cdot, \cdot, \cdot)$. Since the graph G itself was chosen uniformly at random from the set $\mathbb{G}(n, k, m)$, this shows that conditioned on (n_p, R^p, m_p), all graphs $G_p \in \mathcal{P}$ with n_p variable nodes, m_p check nodes, and check degree profile R^p are equally likely to be observed. $\quad\square$

6.2. *Proof of Lemma* 3.12: *Periphery is exponentially peelable.* Let $G = (F, V, E)$ be a graph drawn uniformly at random from $\mathbb{G}(n, k, \alpha n)$, and let $G_P = (F_P, V_P, E_P)$ be its periphery. Recall the connection between BP_0 and the peeling algorithm from Section 4. Let Q be defined as in Theorem 1, that is, Q is the largest positive solution of $Q = 1 - \exp\{-k\alpha Q^{k-1}\}$. In light of Lemma 4.8, we define the asymptotic degree profile pair of the periphery, $(\bar{\alpha}, \bar{R}(x))$ as follows (recall that, from Lemma 4.3, the periphery does include check nodes receiving at most $k - 2$ messages of type 0).

DEFINITION 6.5.

(44) $$\bar{R}(x) \equiv \frac{1}{1 - Q^k - k(1 - Q)Q^{k-1}} \cdot \sum_{l=2}^{k} \binom{k}{l} (1 - Q)^l Q^{k-l} x^l,$$

(45) $$\bar{\alpha} \equiv \alpha \left(\frac{1 - Q^k - k(1 - Q)Q^{k-1}}{1 - Q} \right).$$

Unlike the backbone where all check nodes are of degree k, the periphery can have check nodes of degrees between 2 and k. Among these, check nodes of degree 2 are of importance to us since they can potentially form long strings. Strings

are particularly unfriendly structures for the peeling algorithm; peeling takes linear time to peel such structures. In the next lemma, we define a parameter θ as a function of Q, which is the estimated branching factor of the subgraph of the periphery induced by check nodes of degree 2. Lemma 6.6 proves that this branching factor is less than one for all $\alpha \in (\alpha_d(k), 1]$.

LEMMA 6.6.   *Let* $\theta \equiv \alpha k(k-1)(1-Q)Q^{k-2}$ *with* Q *as defined in Theorem* 1. *Then* $\theta < 1$ *for all* $\alpha \in (\alpha_d(k), 1]$.

Proof of this lemma can be found in the Appendix C.

LEMMA 6.7.   *Let* Q *be defined as in Theorem* 1. *Then there exists* $\eta_1 = \eta_1(\alpha, k) > 0$ *such that the pair* $(\bar{\alpha}, \bar{R})$ *defined in Definition* 6.5 *is peelable at rate* η_1. *Further,* $0 \le f(z, \bar{\alpha}, \bar{R}) \le (1 - \eta_1)z$ *for all* $z \in (0, 1]$.

PROOF.   In view of the density evolution recursion (Definition 14), define

$$f(z) = 1 - \exp(-\bar{\alpha}\bar{R}'(z)).$$

We prove the lemma by showing that $f'(0) = \theta < 1$ and that $f(z) < z$ strictly for $z \in (0, 1]$.

Using the definitions of $\bar{\alpha}$ and $\bar{R}(z)$, the function $f(z)$ can be written as

$$(46) \qquad f(z) = 1 - \exp(-\alpha k((Q + (1-Q)z)^{k-1} - Q^{k-1})).$$

By a straightforward calculation, and using Lemma 6.6, we get

$$(47) \qquad f'(0) = \bar{\alpha}\bar{R}'(0)\exp(-\bar{\alpha}\bar{R}'(0)) = \alpha k(k-1)(1-Q)Q^{k-2} = \theta < 1.$$

Assume $0 \le y \le 1$ to be fixed point of f, that is,

$$(48) \qquad y = 1 - \exp(-\alpha k((Q + (1-Q)y)^{k-1} - Q^{k-1})).$$

Using the identity $Q = 1 - \exp(-\alpha k Q^{k-1})$ and after some calculation, we get

$$(49) \qquad Q + (1-Q)y = 1 - \exp(-\alpha k(Q + (1-Q)y)^{k-1}).$$

Equation (49) shows that $Q + (1-Q)y$ is a fixed point of the original density evolution recursion (14) with $R(x) = x^k$. Since, by definition, Q is the largest fixed point of that recursion, $y = 0$ is the only fixed point of $f(z) = 1 - \exp(-\bar{\alpha}\bar{R}'(z))$ in the interval $[0, 1]$. Since $f'(0) < 1$, we have $f(z) < z$ for all $z \in (0, 1]$ and, therefore, $f(z)/z < 1$ for all $z \in [0, 1]$. The claim follows by taking $\eta_1 = 1 - \sup_{z \in [0,1]} f(z)/z$, with $\eta_1 > 0$ by continuity of $z \mapsto f(z)/z$ over the compact $[0, 1]$. $\square$

We can now prove Lemma 3.12.

PROOF OF LEMMA 3.12.   For any $\varepsilon > 0$, by Lemmas 4.3 and 4.8, we know that

(50)
$$|\alpha_P - \bar{\alpha}| < \varepsilon,$$
$$\left|R_l^P - \bar{R}_l\right| < \varepsilon \qquad \text{for } l \in \{2, \ldots, k\},$$

hold w.h.p.

As before, let $f(z, \alpha, R) = 1 - \exp\{-\alpha R'(z)\}$. Using $R_0^P = R_1^P = 0$ we obtain that the function $f(z, \alpha, R)/z$ is an analytic function over set $[0, 1]^{k+2}$. By Lemma 6.7, $f(z, \bar{\alpha}, \bar{R})/z \le 1 - \eta_1$. It follows that, for $\varepsilon > 0$ small enough, $\partial f(z, \bar{\alpha}, \bar{R})/\partial z \le 1 - (\eta_1/2)$ using continuity $\partial f/\partial z$ with respect to the other arguments of f. We infer that the periphery is w.h.p. peelable at rate $\eta = \eta_1/2$. This proves part (i). Part (ii) follows immediately from Lemma 4.8.   □

7. Proof of Lemma 3.5.   We find it convenient to work within the configuration model: we assume here that G is drawn uniformly at random from $\mathbb{C}(n, k, m)$. The following fact is an immediate consequence of Lemma 5.3.

FACT 7.1.   Assume G is drawn uniformly at random from $\mathbb{C}(n, k, m)$, and denote by n_C, m_C the number of variable and check nodes in the core of G. Suppose $(n_C = n_c, m_C = m_c)$ occurs with positive probability. Then conditioned on $(n_C = n_c, m_C = m_c)$, the core is drawn uniformly from $\mathbb{C}(n_c, k, m_c; 0, n_c)$ [recall the definition of this ensemble in equation (33)].

In words, the core is drawn uniformly from $\mathbb{C}(n_C, k, m_C)$ conditioned on all variable nodes having degree 2 or more.

Now, it has been proved [14] that, w.h.p.

(51) $$\left|n_C/n - \left(1 - \exp(-\alpha k \widehat{Q})(1 + \alpha k \widehat{Q})\right)\right| = o(1),$$

(52) $$\left|m_C/n - \alpha Q^k\right| = o(1),$$

where $(Q, \widehat{Q})$ is as defined in Theorem 1. The above bounds also follow from Lemmas 4.3 and 4.5.

The kernel of the core system $\mathcal{S}_C$ contains all vectors $\underline{x}$ with the following property. Let $V_{(1)} \subseteq V_C$ be the subset of variables taking value 1 in $\underline{x}$ (i.e., the support of $\underline{x}$). Then the subgraph of G_C induced by $V_{(1)}$ has no check node with odd degree.

We will refer to such subgraphs as to *even* subgraphs. Explicitly, even subgraphs are variable-induced subgraphs such that no check node has odd degree. We want characterize the even subgraphs of G_C having no more than $n\varepsilon$ variable nodes, in terms of their size and number. Lemma 7.4 in Section 7.1 below allows us to do this provided certain conditions are met. Our next lemma tells us that the core meets these conditions w.h.p.

LEMMA 7.2. *Fix k and consider any $\alpha \in (\alpha_d(k), \alpha_s(k))$. There exists $\delta = \delta(\alpha, k) > 0$ such that the following happens. Let G be drawn uniformly from $\mathbb{C}(n, k, \alpha n)$. Let n_C be the (random) number of variable nodes in the core, m_C be the number of check nodes in the core and $\alpha_C \equiv m_C/n_C$. Let η_C be the unique positive solution of*

$$(53) \qquad \frac{\eta_C(e^{\eta_C} - 1)}{e^{\eta_C} - 1 - \eta_C} = \alpha_C k$$

and let $\theta_{2C} \equiv \eta_C(k - 1)/(e^{\eta_C} - 1)$. For any $\delta' > 0$, we have, w.h.p.:

   (i) $\theta_{2C} \leq 1 - \delta$.
   (ii) $\alpha_C \in [2/k + \delta, 1]$.
   (iii) $n_C/n \geq (1 - \exp(-\alpha k \widehat{Q})(1 + \alpha k \widehat{Q})) - \delta'$.

The discussion in Section 7.1 throws light on the definitions of η_C and θ_{2C} used.

PROOF OF LEMMA 7.2.   From equations (51), (52), we deduce that $\eta_C = \alpha k \widehat{Q} + o(1)$ w.h.p., leading to

$$\theta_{2C} = \alpha k(k - 1) Q^{k-2}(1 - Q) + o(1) \leq 1 - \delta$$

for sufficiently small δ, using Lemma 6.6. Thus, we have established point (i).

Point (iii) and the lower bound in point (ii) are easy consequences of equations (51), (52). The upper bound in point (ii), $\alpha_C \leq 1$ w.h.p., follows directly from the fact that for $\alpha < \alpha_s$, the system $\mathbb{H}x = \underline{b}$ has a solution for all $\underline{b} \in \{0, 1\}^m$ w.h.p. □

PROOF OF LEMMA 3.5.   Consider first $G \sim \mathbb{C}(n, k, m)$. Applying Fact 7.1 and Lemma 7.2, we deduce that, conditional on the number of nodes, the core is $G_C \sim \mathbb{C}(n_c, k, m_c; 0, n_c)$ and satisfies the conditions of Lemma 7.4 proved below. By Lemma 7.4, the elements of $\mathcal{L}_C(\varepsilon n)$ are in correspondence with simple loops in the subgraph of G_C induced by degree-2 variable nodes. The sparsity bounds follows from Lemma 7.4. The clam that they are, with high probability, disjoint, follows instead from the fact that this random subgraph is subcritical (since $2\alpha R_2 < 1$), and hence decomposes in trees and unicyclic components.

Using Lemma 4.4, we deduce that the result holds also for the $G \sim \mathbb{G}(n, k, m)$ as required. □

7.1. *Characterizing even subgraphs of the core.*   This section aims at characterizing the small even subgraphs of the core G_C. For the sake of simplicity, we shall drop the subscript C throughout the subsection.

Fix k. Consider some $\alpha > 2/k$. Let $\eta_* > 0$ be defined implicitly by

$$(54) \qquad \frac{\eta_*(e^{\eta_*} - 1)}{e^{\eta_*} - 1 - \eta_*} = \alpha k.$$

For $\alpha \in (2/k, \infty)$, we have $\eta_*(\alpha) > 0$ and η_* is an increasing function of α at fixed k [14].

Consider a graph $G = (F, V, E)$ drawn uniformly at random from $\mathbb{C}(n, k, \alpha n; 0, n)$. The rationale for this definition of η_* is that the asymptotic degree distribution of variable nodes in G is Poisson(η_*) conditioned on the outcome being greater than or equal to 2 [to be denoted below Poisson$_{\geq 2}(\eta_*)$].

We are interested in even subgraphs of G.

Consider the subgraph $G_2 = (F, V^{(2)}, E^{(2)})$ of G induced by *variable* nodes of degree 2 (with all factor nodes retained). The asymptotic branching factor this subgraph turns out to be $\theta_2 \equiv \eta_*(k - 1)/(e^{\eta_*} - 1)$. We impose the condition $\theta_2 \leq 1 - \delta$ for some $\delta > 0$ (since this is true of the core). Note that θ_2 is a decreasing function of η_*, and hence a decreasing function of α, for fixed k.

First, we state a technical lemma that we find useful.

LEMMA 7.3. *Consider any k, any $\alpha \in (2/k, 1]$ and $\varepsilon \in (0, 1]$. Then there exists $N_0 \equiv N_0(k, \varepsilon) < \infty$ and $C = C(k) < \infty$ such that the following occurs for all $n > N_0$. Consider a graph $G = (F, V, E)$ drawn uniformly at random from $\mathbb{C}(n, k, m; 0, n)$, $m = n\alpha$. With probability at least $1 - 1/n$, there is no subset of variable nodes $V' \subseteq V$ such that $|V'| \leq \varepsilon n$ and the sum of the degrees of nodes in V' exceeds $C\varepsilon \log(1/\varepsilon)n$.*

PROOF. Let $\deg(i)$ be the degree of variable node $i \in V$. Let $X_i \sim$ Poisson$_{\geq 2}(\eta_*)$ be i.i.d. for $i \in V$. Then $(\deg(i))_{i=1}^n$ is distributed as $(X_i)_{i=1}^n$, conditioned on $\sum_{i=1}^n X_i = mk$. Consider $V' = \{1, 2, \ldots, l\}$. We have

$$\mathbb{P}\left\{\sum_{i=1}^l \deg(i) \geq \gamma l\right\} = \mathbb{P}\left\{\sum_{i=1}^l X_i \geq \gamma l \,\middle|\, \sum_{i=1}^n X_i = mk\right\}$$

$$\leq \frac{\mathbb{P}\{\sum_{i=1}^l X_i \geq \gamma l\}}{\mathbb{P}\{\sum_{i=1}^n X_i = mk\}}.$$

Now, $n\mathbb{E}[X_i] = n\alpha k = mk$, by our choice of η_* in equation (54). Since $\alpha \leq 1$, we deduce that $\eta_* \leq C_1 = C_1(k) < \infty$. Using a local central limit theorem (CLT) for lattice random variables (Theorem 5.4 of [22]) we obtain $\mathbb{P}\{\sum_{i=1}^n X_i = mk\} \geq C_2 n^{-1/2}$ for some $C_2 = C_2(k) > 0$. A standard Chernoff bound yields $\mathbb{P}\{\sum_{i=1}^l X_i \geq \gamma l\} \leq \exp\{-l\gamma C_3\}$, for some $C_3(k) \in (0, 1]$, provided $\gamma > 2\alpha k$. Thus, we obtain

(55) $$\mathbb{P}\left\{\sum_{i=1}^l \deg(i) \geq \gamma l\right\} \leq n^{1/2} \exp\{-l\gamma C_3\}/C_2,$$

provided $\gamma > 2\alpha k$. We use $\gamma = C'(1 + \log(1/\varepsilon))$ with $C' = 2\alpha k/C_3$. Take $l = \varepsilon n$. The number of different subsets of variable nodes of size l is $\binom{n}{l} \leq (e/\varepsilon)^l$ for $n \geq N_1$ for some $N_1 = N_1(\varepsilon) < \infty$. A union bound gives the desired result.   □

LEMMA 7.4.   *Fix $k \geq 3$, and $\delta > 0$ so that for any $\alpha \in [2/k + \delta, 1]$, we have $\theta_2(\alpha, k) \leq 1 - \delta$. Then, for any $\delta' > 0$, there exists $\varepsilon = \varepsilon(\delta, k) > 0$, $C = C(\delta, \delta', k) < \infty$ and $N_0 = N_0(\delta, \delta', k) < \infty$ such that the following occurs for every $n > N_0$. Consider a graph $G = (F, V, E)$ drawn uniformly at random from $\mathbb{C}(n, k, \alpha n; 0, n)$. With probability at least $1 - \delta'$, both the following hold:*

(i) *Consider minimal even subgraphs consisting of only degree 2 variable nodes. There are no more than C such subgraphs. Each of them is a simple cycle consisting of no more than C variable nodes.*

(ii) *Every even subgraph of G with less than εn variable nodes contains only degree 2 variable nodes.*

PROOF.   *Part* (i): Reveal the mk edges of G sequentially. The expected number of nodes in $V^{(2)}$, conditioned on the first t edges revealed forms a martingale with differences bounded by 2. Then, from Azuma–Hoeffding inequality [19], we deduce that $|V^{(2)}|$ concentrates around its expectation:

$$\mathbb{P}(||V^{(2)}| - \mathbb{E}[|V^{(2)}|]| \geq \zeta \sqrt{n}) \leq \exp(-\widehat{C}_1 \zeta^2)$$

for all $\zeta > 0$, where $\widehat{C}_1 = \widehat{C}_1(k) > 0$. The expectation can be computed for instance using the Poisson representation as in the proof of Lemma 7.3, yielding $|\mathbb{E}|V^{(2)}| - n\eta_*^2/(2(e^{\eta_*} - 1 - \eta_*))| \leq n^{3/4}$, for all $\alpha < 1$, $n \geq \widehat{N}_0(k)$. We deduce that for any $\delta_1 = \delta_1(\delta, k) > 0$, we have

$$(56) \qquad \mathbb{P}(||V^{(2)}|/n - \eta_*^2/(2(e^{\eta_*} - 1 - \eta_*))| \geq \delta_1 n) \leq 1/n$$

for all $n > \widehat{N}_1$, where $\widehat{N}_1 = \widehat{N}_1(\delta, k) < \infty$.

Now, condition on $|V^{(2)}| = n^{(2)}$, for some $n^{(2)}$ such that

$$(57) \qquad |n^{(2)}/n - \eta_*^2/(2(e^{\eta_*} - 1 - \eta_*))| < \delta_1 n.$$

Note that by choosing δ_1 small enough, we can ensure $n^{(2)} = \Omega(n)$. We are now interested in the check degree distribution $R^{(2)}$ in G_2. Reveal the $2n^{(2)}$ edges of G_2 sequentially. Consider $l \in \{0, 1, \ldots, k\}$. The expected number of check nodes with degree l in G_2, conditioned on the edges revealed thus far, forms a martingale with differences bounded by 2. Let $Z \sim \text{Binom}(k, 2n^{(2)}/(mk))$. We have $\mathbb{E}[R_l^{(2)}] = \mathbb{P}(Z = l) + O(1/n)$. Arguing as above for each $l \leq k$, we finally obtain

$$(58) \qquad \mathbb{P}\left(\sum_{l=0}^{k} |R_l^{(2)} - \mathbb{P}(Z = l)| \geq \delta_1 n\right) \leq 1/n$$

for all $n > \widehat{N}_2$, where $\widehat{N}_2 = \widehat{N}_2(\delta, k) < \infty$.

Now condition on both $n^{(2)}$ satisfying equation (57) and $R^{(2)}$ satisfying

$$\sum_{l=0}^{k} |R_l^{(2)} - \mathbb{P}(Z = l)| < \delta_1.$$

Let ζ be the branching factor of G_2 (i.e., of a graph that is uniformly random conditional on the degree profile $R^{(2)}$). Under the above conditions on $n^{(2)}$ and $R^{(2)}$, a straightforward calculation implies that ζ is bounded above by $\theta_2 + \delta_2$, for some $\delta_2 = \delta_2(\delta_1, k)$ such that $\delta_2 \to 0$ as $\delta_1 \to 0$. Thus, by selecting appropriately small δ_1, we can ensure that $\delta_2 \leq \delta/2$, leading to a bound of $1 - \delta/2$ on the branching factor for all $n^{(2)}$, $R^{(2)}$ within the range specified above.

Now we condition also on the degree sequence, that is, the sequence of check node degrees in G_2. The factor graph G_2 can be naturally associated to a graph, by replacing each variable node by an edge and each check node by a vertex. This graph is distributed according to the standard (nonbipartite) configuration model. Using [37], Theorem 4, we obtain that the number of cycles of length $l \in \{1, 2, \ldots, l_0\}$ for a constant l_0 are asymptotically independent Poisson random variables, with parameters[8]

$$\lambda_l = \zeta^l/(2l) \qquad \text{for } \zeta = \left[\sum_{d=1}^{k} d(d-1)R^{(2)}(d)\right] \Big/ \left[\sum_{d=1}^{k} dR^{(2)}(d)\right].$$

More precisely, for any constants $c_1, c_2, \ldots, c_{l_0} \in \mathcal{N} \cup \{0\}$, we have

$$\mathbb{P}[\mathsf{E}_n(\underline{c})] = \prod_{l=1}^{l_0} \mathbb{P}(\text{Poisson}(\lambda_l) = c_l) + o(1),$$

where $\mathsf{E}_n(\underline{c})$ is the event that there are c_l cycles of length l for $l \in \{1, 2, \ldots, l_0\}$ with all cycles disjoint from each other, and $\underline{c} = (c_l)_{l=1}^{l_0}$. Choosing l_0 large enough, we have

$$\sum_{\underline{c} \in \mathcal{N}} \mathbb{P}[\mathsf{E}_n(\underline{c})] \geq 1 - \exp\left(-\sum_{l=1}^{\infty} \lambda_l\right) - \delta/4 = 1 - (1 - \zeta)^{-1/2} - \delta'/4,$$

where $\mathcal{N} = \{\underline{c} : \underline{c} \neq \underline{0}, c_l \leq l_0 \text{ for } l \in \{1, 2, \ldots, l_0\}\}$, for n large enough.

On the other hand, we know that the probability of having no cycles in G_2 is $(1 - \zeta)^{-1/2} + o(1)$ under our assumption of $\zeta \leq 1 - \delta/2$. The argument for this was already outlined in the proof of Lemma 3.11, cf. Section 5.1: the Poisson approximation of [37] is used to estimate the probability of having no cycles of length smaller than M, while a simple first moment bound is sufficient for cycles of length M or larger. Thus, with probability at least $1 - \delta'/3$, we have no more than l_0^2 cycles, disjoint and each of length no more than l_0. Choosing $C = l_0^2$, we obtain part (i) with probability at least $1 - \delta'/2$ for large enough n.

Part (ii): Let $m \equiv \alpha n$. Let $\mathcal{N}(G; l, j)$ be the number of even subgraphs of G induced by l variable nodes such that the sum of the degrees of the l variable

nodes is $2(l + j)$. We are interested in $l \leq \varepsilon n$ (we will choose ε later) and $j > 0$. In particular, we want to show that, for any $\delta' > 0$,

$$(59) \qquad \mathbb{P}\left\{\sum_{l=1}^{\varepsilon n} \sum_{j=1}^{mk/2} \mathcal{N}(G; l, j) > 0\right\} \leq \delta'/2.$$

This immediately implies the desired result from linearity of expectation and Markov inequality.

From Lemma 7.3, we deduce that

$$(60) \qquad \mathbb{P}\left\{\sum_{l=1}^{\varepsilon n} \sum_{j=\varepsilon'n}^{mk/2} \mathcal{N}(G; l, j) > 0\right\} \leq 1/n,$$

for some $\varepsilon'(\varepsilon, k)$ with the property that $\varepsilon' \to 0$ as $\varepsilon \to 0$. Thus, we only need to establish

$$(61) \qquad \sum_{l=1}^{\varepsilon n} \sum_{j=1}^{\varepsilon'n} \mathbb{E}[\mathcal{N}(G; l, j)] \leq \delta'/3,$$

for all n large enough, since the claim then follows from Markov inequality.

A straightforward calculation [29, 36] yields

$$\mathbb{E}[\mathcal{N}(G; l, j)] = \frac{\binom{n}{l}\mathcal{T}_1\mathcal{T}_2\mathcal{T}_3}{\binom{mk}{2(l+j)}\mathcal{T}_4},$$

where

$$\mathcal{T}_1 = \text{coeff}\big[(e^y - 1 - y)^l; y^{2(l+j)}\big],$$
$$\mathcal{T}_2 = \text{coeff}\big[(e^y - 1 - y)^{n-l}; y^{mk-2(l+j)}\big],$$
$$\mathcal{T}_3 = \text{coeff}\left[\left(\frac{(1+y)^k + (1-y)^k}{2}\right)^m; y^{2(l+j)}\right],$$
$$\mathcal{T}_4 = \text{coeff}\big[(e^y - 1 - y)^n; y^{mk}\big].$$

It is useful to recall the following probabilistic representation of combinatorial coefficients.

FACT 7.5.   For any $\eta > 0$, we have

$$(62) \qquad \text{coeff}\big[(e^y - 1 - y)^N; y^M\big] = \eta^{-M}(e^\eta - 1 - \eta)^N \mathbb{P}\left[\sum_{i=1}^N X_i = M\right],$$

where $X_i \sim \text{Poisson}_{\geq 2}(\eta)$ are i.i.d. for $i \in \{1, \ldots, M\}$.

Consider $\mathcal{T}_4$. By definition, cf. equation (54), η_* is such that for $X_i \sim$ Poisson$_{\geq 2}(\eta_*)$ we have $\mathbb{E}[X_i] = \alpha k = mk/n$. Moreover, $\alpha \in [2/k + \delta, 1]$ implies $\eta_* \in [C_1, C_2]$ for some $C_1 = C_1(\delta, k) > 0$ and $C_2 = C_2(k) < \infty$. From $\eta_* \leq C_2$ and using a local CLT for lattice random variables [22], it follows that $\mathbb{P}[\sum_{i=1}^n X_i = mk] \geq C_3/\sqrt{n}$ for some $C_3 = C_3(\delta, k) > 0$. Thus, using Fact 7.5, we have

$$\tag{63} \mathcal{T}_4 \geq \eta_*^{-mk} (e^{\eta_*} - 1 - \eta_*)^n C_3 n^{-1/2}.$$

Now, consider $\mathcal{T}_2$. Again use $\eta = \eta_*$ in Fact 7.5. From $\eta_* \geq C_1$ and again using a local CLT for lattice r.v.'s [22], we obtain $\mathbb{P}[\sum_{i=1}^{n-l} X_i = mk - 2(l + j)] \leq C_4/2\sqrt{n-l} \leq C_4/\sqrt{n}$ for some $C_4 = C_4(\delta, k) < \infty$, since $l \leq \varepsilon n$. Thus, Fact 7.5 yields

$$\tag{64} \mathcal{T}_2 \leq \eta_*^{-mk+2(l+j)} (e^{\eta_*} - 1 - \eta_*)^{n-l} C_4 n^{-1/2}.$$

Fact 7.5 yields that $\mathcal{T}_1$ can be bounded above as

$$\tag{65} \mathcal{T}_1 \leq \eta^{-2(l+j)} (e^{\eta} - 1 - \eta)^l$$

for any $\eta > 0$. We will choose a suitable η later.

Finally, for $\mathcal{T}_3$, similar to Fact 7.5, we can deduce that

$$\mathcal{T}_3 \leq \left(\frac{(1 + \xi)^k + (1 - \xi)^k}{2} \right)^m \xi^{-2(l+j)}$$

for all $\xi > 0$. Now, it is easy to check that

$$\frac{(1 + \xi)^k + (1 - \xi)^k}{2} \leq \exp\left\{ \binom{k}{2} \xi^2 \right\},$$

by comparing coefficients in the series expansions of both sides. Choosing $\xi = \sqrt{(l + j)/(m\binom{k}{2})}$, we obtain

$$\tag{66} \mathcal{T}_3 \leq \left(\frac{em\binom{k}{2}}{l + j} \right)^{l+j}.$$

Finally, we have

$$\tag{67} \binom{n}{l} \leq \frac{n^l}{l!}, \qquad \binom{mk}{2(l+j)} \geq \frac{(mk - 2(l+j))^{2(l+j)}}{(2(l+j))!}.$$

Putting together equations (63), (64), (65), (66) and (67), we obtain

$$\mathbb{E}[\mathcal{N}(G; l, j)] \leq C_6 \cdot \frac{(e^{\eta} - 1 - \eta)^l}{\eta^{2(l+j)}} \cdot \frac{\eta_*^{2(l+j)}}{(e^{\eta_*} - 1 - \eta_*)^l}$$

$$\times \left(\frac{e(k-1)(1 + C_5((l+j)/n))}{2(l+j)k} \right)^{l+j} \cdot \frac{(2(l+j))!}{l! \alpha^l m^j},$$

for some $C_6 = C_6(k, \delta) < \infty$. Now, $N! \geq C_7 \sqrt{N}(N/e)^N$ for all $N \in \mathbb{N}$, for some $C_7 > 0$. Using this with $N = l + j$, we obtain

$$
\left(\frac{e}{l+j}\right)^{l+j} \cdot \frac{(2(l+j))!}{l!} \leq \frac{\sqrt{l+j}}{C_7} \cdot \frac{(2(l+j))!}{l!(l+j)!}
$$
$$
\leq \frac{\sqrt{l+j}}{C_7 l^j} \cdot \binom{2(l+j)}{(l+j)} \leq \frac{C_8 2^{2(l+j)}}{l^j},
$$

for some $C_8 < \infty$. Plugging back, we get

$$
\mathbb{E}[\mathcal{N}(G; l, j)] \leq C_9 (\mathcal{T}_5)^l (\mathcal{T}_6)^j,
$$

where

$$
\mathcal{T}_5 = 2\theta_2 \frac{(e^\eta - 1 - \eta)}{\eta^2}(1 + C_5((l+j)/n)),
$$

$$
\mathcal{T}_6 = \frac{4(k-1)\eta_*^2}{ml\eta^2}.
$$

Without loss of generality, assume $\delta \leq 0.1$. Now, we choose $\varepsilon = \varepsilon(\delta, k) > 0$ such that $\varepsilon + \varepsilon' \leq \delta/(10C_5)$. We choose $\eta = \eta(k) > 0$ such that $(e^\eta - 1 - \eta)\eta^{-2} \leq (1 + \delta/10)/2$ [note that $(e^\eta - 1 - \eta)\eta^{-2} \to 1/2$ as $\eta \to 0$]. This leads to $\mathcal{T}_5 \leq 1 - \delta/2$ for all $l \leq \varepsilon n$ and $j \leq \varepsilon' n$, when we use $\theta_2 \leq 1 - \delta$. Also, $\mathcal{T}_6 \leq C_{10}/n$ for all l, j, for some $C_{10} = C_{10}(k) < \infty$. Thus,

$$
\mathbb{E}[\mathcal{N}(G; l, j)] \leq C_9 (1 - \delta/2)^l \left(\frac{C_{10}}{n}\right)^j.
$$

Summing over j and l, we obtain

$$
(68) \qquad \sum_{l=1}^{\varepsilon n} \sum_{j=1}^{\varepsilon' n} \mathbb{E}[\mathcal{N}(G; l, j)] \leq \frac{C_{11}}{n}
$$

for some $C_{11} = C_{11}(k, \delta) < \infty$. This implies equation (61) for large enough n as required. $\square$

8. Proof of Lemma 3.8: A sparse basis for low-weight core solutions.

For each $\underline{x}_C \in \mathcal{L}_C(\varepsilon n)$, we need to find a sparse solution $\underline{x} \in \mathcal{S}_1$ that matches $\underline{x}_C$ on the core. From Lemma 3.5, we know that w.h.p., $\underline{x}_C$ consists of all zeros except for a small subset of variables. Indeed, we know from Lemma 7.4 that these variables correspond to a cycle of degree-2 variable nodes. Although this is not used in the following, we shall nevertheless refer to the set of variable nodes corresponding to an element of $\mathcal{L}_C(\varepsilon n)$ as a cycle. Denote by L_1 the cycle corresponding to $\underline{x}_C$. Recall that the noncore $G_{NC} = (F_{NC}, V_{NC}, E_{NC})$ is the subgraph of G induced by $F_{NC} = F \setminus F_C$ and $V_{NC} = V \setminus V_C$. Suppose we set all noncore variables to 0. The set of violated checks consists of those checks in F_{NC} that have an odd number of

neighbors in L_1. We show that w.h.p., each such check can be satisfied by changing a small number of noncore variables in its neighborhood to 1. To show that this is possible, we make use of the belief propagation algorithm described in Section 4.

Our strategy is roughly the following. Consider a violated check a. We wish to set an odd number of its noncore neighboring variables to 1. But then, this may cause further checks to be violated, and so on. A key fact comes to our rescue. If check node a receives an incoming $*$ message in round T, then we can find a subset of noncore variable nodes in a T-neighborhood of a such that if we set those variables to 1, check a will be satisfied (with an odd number of neighboring ones in the noncore) without causing any new violations. We do this for each violated check. Now w.h.p., for suitable T, all violated checks will receive at least one incoming $*$ by time T (note that each noncore check receives an incoming $*$ at the BP fixed point). Thus, we can satisfy them all by setting a small number of noncore variables to 1.

LEMMA 8.1.   *Consider G drawn uniformly from $\mathbb{G}(n, k, m)$. Denote by $F^{(l)} \subseteq F_{\mathrm{NC}}$ the checks in the noncore having degree l with respect to the noncore, for $l \in \{1, 2, \ldots, k\}$. Condition on the core G_{C}, and $F^{(l)}$ for $l \in \{1, 2, \ldots, k\}$.*

- *Then $E_{\mathrm{C,NC}}$ and G_{NC} are independent of each other. Here $E_{\mathrm{C,NC}}$ denotes the edges between core variables V_{C} and noncore checks F_{NC}.*
- *The edges in $E_{\mathrm{C,NC}}$ are distributed as follows: For each $a \in F_{\mathrm{NC}}$, if $a \in F^{(l)}$, its neighborhood in G_{C} is a uniformly random subset of V_{C} of size $k - l$, independent of the others.*
- *Clearly, $(G_{\mathrm{C}}, (F^{(l)})_{l=1}^k)$ uniquely determine the parameters $(n_{\mathrm{NC}}, R^{\mathrm{NC}}, m_{\mathrm{NC}})$ of the noncore. The noncore G_{NC} is drawn uniformly at random from $\mathbb{D}(n_{\mathrm{NC}}, R^{\mathrm{NC}}, m_{\mathrm{NC}})$ conditioned on being peelable, that is, G_{NC} is drawn uniformly at random from $\mathbb{D}(n_{\mathrm{NC}}, R^{\mathrm{NC}}, m_{\mathrm{NC}}) \cap \mathcal{P}$.*

PROOF.   Each $G \in \mathbb{G}(n, k, m)$ with the given $(G_{\mathrm{C}}, (F^{(l)})_{l=1}^k)$ has a G_{NC} corresponding to a unique element of $\mathbb{D}(n_{\mathrm{NC}}, R^{\mathrm{NC}}, m_{\mathrm{NC}}) \cap \mathcal{P}$ and $E_{\mathrm{C,NC}}$ corresponding to a subset of V_{C} of size $k - l$ for each $a \in F^{(l)}$, for $l \in \{1, \ldots, k\}$. The converse is also true. This yields the result.   □

PROOF OF LEMMA 3.8.   Take any sequence $(s_n)_{n \geq 1}$ such that $\lim_{n \to \infty} s_n = \infty$ and $s_n \leq \varepsilon n$. If points (i), (ii) and (iii) in Lemma 3.5 hold, let V_{cycle} denote the union of the supports of the solutions in $\mathcal{L}_{\mathrm{C}}(s_n)$. Let

$$\mathsf{E}_1 \equiv \mathsf{E}_{1,a} \cap \mathsf{E}_{1,b} \cap \mathsf{E}_{1,c},$$

$$\mathsf{E}_{1,a} \equiv \{\text{Points (i), (ii) and (iii) in Lemma 3.5 hold}\},$$

$$\mathsf{E}_{1,b} \equiv \{|F^{(l)}| \geq n/C_2 \text{ for all } l \in \{1, 2, \ldots, k\}\},$$

$$\mathsf{E}_{1,c} \equiv \{\text{No variable in } V_{\mathrm{cycle}} \text{ has degree exceeding } \log s_n\}.$$

(Note that these events are implicitly indexed by n.) We argue that E_1 holds w.h.p. for an appropriate choice of $C_2 = C_2(k, \alpha) < \infty$. Indeed, Lemma 3.5 implies that $\mathsf{E}_{1,a}$ holds w.h.p. Lemma 4.8 implies that $\mathsf{E}_{1,b}$ holds w.h.p. for sufficiently large C_2. Finally, Lemma 8.1 and a subexponential tail bound on the Poisson distribution ensure $\mathsf{E}_{1,c}$ holds w.h.p.

Assume that E_1 holds. Let sets of variable nodes on the disjoint cycles corresponding to elements of $\mathcal{L}_\mathrm{C}(\varepsilon n)$ be denoted by L_i for $i \in \{1, 2, \ldots, |\mathcal{L}_\mathrm{C}(\varepsilon n)|\}$. Consider a cycle L_i. Denote by a_{ij}, $j \in \{1, 2, \ldots, Z_i\}$, the checks in the noncore having an odd number of neighbors in L_i. (Thus, Z_i is the number of such checks.) Call these *marked* checks. Given E_1, we know that $Z_i \leq s_n \log s_n$, and that there are no more than $s_n^2 \log s_n$ marked checks in total:

$$\sum_{i=1}^{|\mathcal{L}_\mathrm{C}(\varepsilon n)|} Z_i \leq s_n^2 \log s_n.$$

Define

$$\mathsf{E}_2 \equiv \{\text{No more than } n/s_n^3 \text{ messages change after } T_n \text{ iterations of } \mathrm{BP}_0\}.$$

By Lemma 4.10, the event E_2 holds w.h.p. provided $\lim_{n \to \infty} T_n = \infty$ and s_n grows sufficiently slowly with n [for the given choice of $(T_n)_{n \geq 1}$].

Let

$$B_{ij} \equiv \{\text{Not all messages incoming to check } a_{ij} \text{ have converged}$$

$$\text{to their fixed-point value in } T_n \text{ iterations}\}.$$

We wish to show that

(69) $$\bigcap_{i,j} B_{ij}^{\mathrm{c}}$$

holds w.h.p. We have

$$\mathbb{P}\left(\bigcup_{i,j} B_{ij}\right) \leq \mathbb{E}_{(G_\mathrm{C}, E_{\mathrm{C,NC}})}\left[\mathbb{E}\left[\mathbb{I}[\mathsf{E}_1, \mathsf{E}_2] \sum_{i,j} \mathbb{I}[B_{ij}] \Big| G_\mathrm{C}, E_{\mathrm{C,NC}}\right]\right] + \mathbb{P}[\mathsf{E}_1^{\mathrm{c}}] + \mathbb{P}[\mathsf{E}_2^{\mathrm{c}}].$$

Given E_2, we know that the number of checks for which an incoming message changes after T_n is no more than n/s_n^3. Suppose $a_{ij} \in F^{(l)}$ is a marked check. Then we have

$$\mathbb{E}\left[\mathbb{I}[\mathsf{E}_1, \mathsf{E}_2]\mathbb{I}[B_{ij}] | G_\mathrm{C}, E_{\mathrm{C,NC}}\right] \leq \frac{n}{s_n^3 |F^{(l)}|} \leq \frac{1}{C_2 s_n^3},$$

since all check nodes in $F^{(l)}$ are equivalent with respect to the noncore, from Lemma 8.1. We already know that under E_1, the number of marked checks is bounded by $s_n^2 \log s_n$. This leads to

$$\mathbb{P}\left(\bigcup_{ij} B_{ij}\right) \leq \frac{\log s_n}{C_2 s_n} + \mathbb{P}[\mathsf{E}_1^{\mathrm{c}}] + \mathbb{P}[\mathsf{E}_2^{\mathrm{c}}] \overset{n \to \infty}{\longrightarrow} 0,$$

implying equation (69) holds w.h.p.

Condition on G_C and $E_{C,NC}$. This identifies the marked checks. Lemma 8.1 guarantees us that all checks in $F^{(l)}$ are equivalent with respect to G_{NC}. Suppose E_1 holds. Define a ball of radius t around a check node as consisting of the neighboring variable nodes, and the balls of radius t around each of those variables. Similar to the proof of Lemma 3.11(iii), we can show that

$$(70) \qquad \left| B_{G_{NC}}(a_{ij}, T_n) \right| \le C_3^{T_n}$$

holds with probability at least $1 - C_4 \exp(-2^{T_n}/C_4)$, for some $C_3 = C_3(\alpha, k) < \infty$ and $C_4 = C_4(\alpha, k) < \infty$, for all marked checks a_{ij}. Thus, the probability that this bound on ball size holds simultaneously for all marked checks, by union bound, is at least $1 - s_n^2 \log s_n C_4 \exp(-2^{T_n}/C_4) \to 1$ as $n \to 1$ provided $T_n \to \infty$ and s_n grows sufficiently slowly with n.

Suppose equation (69) and E_1 hold. Consider any marked check a_{ij} adjacent to $v \in L_i$ for any L_i. It receives at least one incoming $*$ message at the BP_0 fixed point and since $B_{ij} = 0$, this is also true after T_n iterations of BP_0. Hence, there is a subset of variables $V^{(ij)} \subseteq B_{G_{NC}}(a_{ij}, T_n)$, such that setting variables in $V^{(ij)}$ to 1 satisfies a_{ij} without violating any other checks. Define

$$V^{(i)} \equiv \left\{ v : v \text{ occurs an odd number of times in the sets } (V^{(ij)})_{j=1}^{Z_i} \right\}.$$

It is not hard to verify that the vector $\underline{x}_{c,i}$ with variables in $L_i \cup V^{(i)}$ set to one and all other variables set to zero, is a member of S_1. If equation (70) holds for all marked checks, then we deduce that $|V^{(i)}| \le C_3^{T_n} s_n \log s_n \le c_n$ for T_n and s_n growing sufficiently slowly with n. Thus, $\underline{x}_{c,i} \in S_1$ is c_n-sparse assuming these events, each of which occurs w.h.p. We repeat this construction for every L_i. □

APPENDIX A: PROOF OF LEMMA 3.4

LEMMA A.1. *Assume that G has no 2-core, and let*

$$\mathbb{K} \equiv \left[\begin{pmatrix} \mathbb{H}_{F,U}^{-1} \mathbb{H}_{F,W} \\ I_{(n-m)\times(n-m)} \end{pmatrix} \right],$$

where U and W are constructed as in Lemma 3.2, we order the variables as U followed by W, and the matrix inverse is taken over $G\mathbb{F}[2]$. Then the columns of $\mathbb{K}$ form a basis of the kernel of S, which is also the kernel of $\mathbb{H}$. In addition, if $\mathbb{K}_{i,j} = 1$, then $d_G(i, j) \le T_C$.

PROOF. A standard linear algebra result shows that $\mathbb{K}$ is a basis for the kernel of $\mathbb{H}$. The bottom identity block of $\mathbb{K}$ corresponds to the $(n-m)$ independent variables $w \in W$, and in this block a 1 only occurs if the row and column correspond to the same variable, that is, for $i, j \in W$, $\mathbb{K}_{i,j} = 1$ implies $i = j$, and thus

$d_G(i, j) = 0$. To prove the distance claim for the upper block of $\mathbb{K}$, we proceed by induction on T_C. For a variable $u \in U$ that is peeled along with factor node $a \in F$, we will reference u via the factor node it was peeled with as u_a.

- *Induction base*: For $T_C = 1$, $\mathbb{H}_{F,U} = I_m$, and thus

$$\mathbb{K} = \left[\begin{pmatrix} \mathbb{H}_{F,W} \\ I_{(n-m) \times (n-m)} \end{pmatrix} \right].$$

  Since $T_C = 1$, note that ever variable node must be connected to no more than 1 factor node. Thus, $(\mathbb{H}_{F,W})_{a,i} = 1$ implies that factor node a was connected to independent variable node i. Thus, variables i and u_a are both adjacent to factor a, and consequently $d_G(u_a, i) = 1$.

- *Inductive step*: Assume that $T_C = T + 1$ and consider the graph $\mathsf{J}(G) = (F_J, V_J, E_J)$ (recall that J denoted the peeling operator). By construction $T_C(\mathsf{J}(G)) = T$, and thus by the inductive hypothesis the columns of

$$\mathbb{K}_{\mathsf{J}(G)} \equiv \left[\begin{matrix} \tilde{\mathbb{K}} \\ I_{((n-n_1)-(m-m_1)) \times ((n-n_1)-(m-m_1))} \end{matrix} \right]$$

$$\equiv \left[\begin{matrix} \mathbb{H}_{F_J,U_J}^{-1} \mathbb{H}_{F_J,W_J} \\ I_{((n-n_1)-(m-m_1)) \times ((n-n_1)-(m-m_1))} \end{matrix} \right],$$

form a basis for the kernel of $\mathbb{H}_{\mathsf{J}(G)}$, where F_J, U_J, and W_J refer to the set of factor nodes of the factor graph $\mathsf{J}(G)$, and their corresponding partition, respectively. In addition, $(\mathbb{K}_{\mathsf{J}(G)})_{a,i} = 1$ only if $d_{\mathsf{J}(G)}(u_a, i) \le T$. To extend this basis to a basis for the kernel of $\mathbb{H}$, note that

$$\mathbb{K} \equiv \left[\begin{matrix} \mathbb{H}_{F,U}^{-1} \mathbb{H}_{F,W} \\ I_{(n-m) \times (n-m)} \end{matrix} \right] = \left[\begin{matrix} \begin{pmatrix} \mathbb{H}_{F_1,U_1} & \mathbb{H}_{F_1,U_J} \\ 0 & \mathbb{H}_{F_J,U_J} \end{pmatrix}^{-1} \begin{pmatrix} \mathbb{H}_{F_1,W_1} & \mathbb{H}_{F_1,W_J} \\ 0 & \mathbb{H}_{F_J,W_J} \end{pmatrix} \\ I_{(n-m) \times (n-m)} \end{matrix} \right]$$

$$= \left[\begin{matrix} \begin{pmatrix} I_{|U_1|} & -\mathbb{H}_{F_1,U_J} \mathbb{H}_{F_J,U_J}^{-1} \\ 0 & \mathbb{H}_{F_J,U_J}^{-1} \end{pmatrix} \begin{pmatrix} \mathbb{H}_{F_1,W_1} & \mathbb{H}_{F_1,W_J} \\ 0 & \mathbb{H}_{F_J,W_J} \end{pmatrix} \\ I_{(n-m) \times (n-m)} \end{matrix} \right]$$

$$= \left[\begin{matrix} \begin{pmatrix} \mathbb{H}_{F_1,W_1} & \mathbb{H}_{F_1,W_J} + \mathbb{H}_{F_1,U_J} \tilde{\mathbb{K}} \\ 0 & \tilde{\mathbb{K}} \end{pmatrix} \\ I_{(n-m) \times (n-m)} \end{matrix} \right].$$

By construction if $(\mathbb{H}_{F_1,W_1})_{a,i} = 1$, then $d_G(u_a, i) = 1 \le T$. Consider the (a, i) entry of the matrix $B \equiv \mathbb{H}_{F_1,W_J} + \mathbb{H}_{F_1,U_J} \tilde{\mathbb{K}}$. A necessary condition for $B_{a,i} = 1$ is the existence of an edge between check node $a \in F_1$ and independent variable node $i \in W \setminus W_1 = W_J$ [i.e., $(\mathbb{H}_{F_1,W_J})_{a,i} = 1$], or the existence of both an edge between $a \in F_1$ and dependent variable node $j \in U_J$ that is in the basis for independent variable i [i.e., $((\mathbb{H}_{F_1,U_J})_{a,j} = 1, \tilde{\mathbb{K}}_{j,i} = 1)$].

We note that if $d_{\mathsf{J}(G)}(u_a, i) \le T$, then $d_G(u_a, i) \le T$ also, since $E_J \subset E$. Thus, if $(\mathbb{H}_{F_1,U_J})_{a,j} = 1$, $\tilde{\mathbb{K}}_{j,i} = 1$, then $d_G(u_a, i) \le T + 1$. Similarly, if

$(\mathbb{H}_{F_1, W_J})_{a,i} = 1$, then $d_G(u_a, i) = 1$ as in the base case. Thus, if $\mathbb{K}_{i,j} = 1$, then $d_G(i, j) \le T + 1 = T_C$. $\square$

A direct result of this is the sparsity bound given below.

LEMMA A.2. *For $\mathbb{K}$ constructed as in Lemma A.1, the columns of $\mathbb{K}$ form an s-sparse basis for the kernel of $\mathbb{H}$, with*

$$s \le \max_{i \in V} |\mathsf{B}_G(i, T_C)|.$$

PROOF. By Lemma A.1, $d_G(a, i) \le T_C$ is a necessary condition for $\mathbb{K}_{a,i} = 1$. Thus, for all $i \in W$, the ith column of $\mathbb{K}$ can only contain 1's on the entries that correspond to variables at distance at most T_C from i. The result follows by taking a union bound over all $i \in W$. $\square$

PROOF OF LEMMA 3.4. Let

$$\widehat{\mathbb{K}} = \mathbb{L} \begin{bmatrix} \mathbb{Q}_{F_*, U_*}^{-1} \mathbb{Q}_{F_*, W_*} \\ I_{(n-m) \times (n-m)} \end{bmatrix},$$

where the matrix inverse is taken over $\mathrm{GF}[2]$. If $G_* \ne G$, then all degree 2 check nodes constrain their adjacent variable nodes to the same value. Therefore, all variables in the same connected component take on the same value in a satisfying solution, that is, for all $v_* \in V_*$, if $\mathbb{H}\underline{x} = 0$, then for all $i \in v_*$, either $x_i = 0$ or $x_i = 1$. Consequently, $\mathbb{H}\underline{x} = 0$ if and only if $\underline{x} = \mathbb{L}\underline{x}_*$ for some $\underline{x}_*$ such that $\mathbb{Q}\underline{x}_* = 0$ Thus, $\{\underline{x}^{(1)}, \ldots, \underline{x}^{(N)}\}$ is a basis for the kernel of $\mathbb{H}$ if and only if $\underline{x}^{(i)} = \mathbb{L}\underline{x}_*^{(i)}$ and $\{\underline{x}_*^{(1)}, \ldots, \underline{x}_*^{(N)}\}$ is a basis for the kernel of $\mathbb{Q}$.

Finally notice that $\mathbb{L}\underline{x}_*$ has $|v_*|$ nonzero entries for each $v_* \in V_*$ such that $\underline{x}_{*, v_*} \ne 0$. Thus, the sparsity bound follows as a direct extension of the bound from Lemma A.2, and the columns of $\widehat{\mathbb{K}}$ form an s-sparse basis for the kernel of $\mathbb{H}$, with

$$s \le \max_{v_* \in V_*} S(v_*, T_C(G_*)). \qquad \square$$

APPENDIX B: PROOFS OF TECHNICAL LEMMAS IN SECTION 5

PROOF OF LEMMA 5.2. Let $\omega \equiv \alpha R'(1)$. Define $f(z) \equiv 1 - \lambda(1 - \rho(z)) = 1 - \exp(-\alpha R'(1)\rho(z))$. We obtain

(71)
$$f'(0) = 2\alpha R_2.$$

Now, we know that $z_t \to 0$ as $t \to \infty$, it follows that $\lim_{t \to \infty} z_{t+1}/z_t \to f'(0)$. We then deduce from peelability at rate η that

(72)
$$f'(0) \le 1 - \eta.$$

Combining equations (71) and (72), we obtain the desired result (i).

In order to prove (ii) notice that, for the pair to be peelable, need $z \leq 1 - \exp(-\alpha R'(z))$ for all $z \in [0, 1]$, that is,

$$(73) \qquad R'^{-1}(x) \leq 1 - e^{-\alpha x} \qquad \text{for all } x \in [0, R'(1)],$$

where R'^{-1} is the inverse mapping of $z \mapsto R'(z)$. We next integrate the above over $[0, R'(1)]$, using

$$(74) \qquad \int_0^{R'(1)} R'^{-1}(x)\,dx = \int_0^1 w R''(w)\,dw = R'(1) - 1,$$

$$(75) \qquad \int_0^{R'(1)} (1 - e^{-\alpha x})\,dx = R'(1) - \frac{1}{\alpha}(1 - e^{-\alpha R'(1)}).$$

We thus obtain

$$(76) \qquad 1 \geq \frac{1}{\alpha}(1 - e^{-\alpha R'(1)}),$$

which yields $\alpha \leq 1 - e^{-\alpha R'(1)} < 1$. $\quad \square$

PROOF OF LEMMA 5.3. We use the notation $(G) = (m_l(G))_{l=2}^k$ whereby $m_l(G)$ is the number of check nodes of degree l in G. Let

$$n_1^{(t)} \equiv n_1(J_t), \qquad n_2^{(t)} \equiv n_2(J_t), \qquad \underline{m}^{(t)} \equiv \underline{m}(J_t),$$

$$\alpha^{(t)} \equiv \left(\sum_{l=2}^k m_l^{(t)}\right) \bigg/ n, \qquad R_l^{(t)} \equiv m_l^{(t)} \bigg/ \left(\sum_{l'=2}^k m_{l'}^{(t)}\right) \qquad \text{for } l \in \{2, 3, \ldots, k\}.$$

Note that $R^{(t)}$ defined above is, in fact, the check degree profile of J_t.

As above, let $J(\cdot)$ denote the operator corresponding to one round of synchronous peeling [so that $J_t = J^t(G)$]. Define the set

$$S(G; \widehat{\underline{m}}, \widehat{n}_1, \widehat{n}_2) \equiv \{\widehat{G} : n_1(\widehat{G}) = \widehat{n}_1, n_2(\widehat{G}) = \widehat{n}_2, \underline{m}(\widehat{G}) = \widehat{\underline{m}}, J(\widehat{G}) = G\}.$$

We prove the result by induction. By definition, we know that $J_0 = G$ is drawn uniformly from the $\mathbb{C}(n, R, \alpha n)$. Suppose, conditioned on $\underline{m}^{(t)}, n_1^{(t)}, n_2^{(t)}$, the graph J_t is drawn uniformly from $\mathbb{C}(n, R^{(t)}, \alpha^{(t)} n)$. Let the probability of each possible J_t [with parameters $(\underline{m}^{(t)}, n_1^{(t)}, n_2^{(t)})$] be denoted by $q(\underline{m}^{(t)}, n_1^{(t)}, n_2^{(t)})$. Consider a candidate graph G' with parameters $(\underline{m}', n_1', n_2')$. We have

$$\mathbb{P}[J_{t+1} = G'] = \sum_{J_t : J(J_t) = G'} \mathbb{P}[J_t]$$

$$= \sum_{\widehat{\underline{m}}, \widehat{n}_1, \widehat{n}_2} \sum_{J_t \in S(G'; \widehat{\underline{m}}, \widehat{n}_1, \widehat{n}_2)} \mathbb{P}[J_t]$$

$$= \sum_{\widehat{\underline{m}}, \widehat{n}_1, \widehat{n}_2} q(\widehat{\underline{m}}, \widehat{n}_1, \widehat{n}_2) |S(G'; \widehat{\underline{m}}, \widehat{n}_1, \widehat{n}_2)|.$$

A straightforward count yields

$$|S(G'; \underline{\widehat{m}}, \widehat{n}_1, \widehat{n}_2)|$$
$$= \binom{n - n'_1 - n'_2}{\widehat{n}_1} \cdot \Delta! \cdot \text{coeff}\big[(e^z - 1)^{n'_1}(e^z)^{n'_2}; z^{\Delta - \widehat{n}_1}\big] \cdot \mathbb{I}[\widehat{n}_2 = n'_1 + n'_2],$$

where $\Delta \equiv \sum_{l=1}^{k}(\widehat{m}_l - m'_l)l$. Thus, $\mathbb{P}[J_{t+1} = G']$ depends on G' only through $(\underline{m}', n'_1, n'_2)$. □

To simplify the proof of Lemma 5.4, we first prove a simple technical lemma.

LEMMA B.1.   *Let $G = (F, V, E)$ be a factor graph that is a tree with no check node of degree 1 or 2, rooted at a variable node v, with $|V| > 1$. Then $|\{u \in V : \deg(u) \le 1, u \ne v\}| \ge |V|/2$, that is, at least half of all variable nodes are leaves. (Here, a leaf is defined as a variable node that is distinct from the root and has degree at most 1.)*

PROOF.   We proceed by induction on the maximum depth t of the tree G rooted at v.

- *Induction base*: For a tree of depth 1, let $c = \deg(v) > 0$. Since all check nodes have degree 3 or more, G has $N_1 \ge 2c$ leaves and $|V| = N_1 + 1$. Clearly, $N_1 \ge |V|/2$.
- *Inductive step*: Consider G having depth $t + 1$ and perform 1 round of synchronous peeling, resulting in $\mathsf{J}(G) = G' = (F', V', E')$. Let N'_1 be the number of leaves in V'. The inductive hypothesis implies $|V'| \le 2N'_1$, since G' is also a tree. Since, by construction, every factor node has degree at least 3 in G, every leaf in G' must have at least 2 leaves in G as descendants, that is, $2N'_1 \le N_1$, where N_1 is the number of leaves in G. Combining these two inequalities yields

$$|V| = |V'| + N_1 \le 2N'_1 + N_1 \le 2N_1,$$

as desired. □

PROOF OF LEMMA 5.4.   By Lemma B.1, if G is a tree, at least one-half of all variable nodes are leaves at every stage of peeling. Thus, G is peelable and $T_C(G) \le \lceil \log_2 |V| \rceil$. (After $\lceil \log_2 |V| \rceil - 1$ rounds of peeling, we have 2 or less variable nodes remaining, and hence no checks. At most one more round of peeling leads to annihilation.)

Now suppose G is unicyclic. Each factor in the cycle has degree at least 3, hence it has a neighbor outside the cycle and must eventually get peeled. Breaking ties arbitrarily, let a be the first factor in the cycle to be peeled, and let $u \in \partial a$ be the variable node that "causes" it to get peeled (clearly u is not in the cycle). Let $t_u \le T_C(G)$ be the peeling round in which u and a are peeled. Consider the subtree

$G_u = (F_u, V_u, E_u)$ rooted at u defined as follows: G_u is the maximal connected subgraph of G that includes u, but not a. Using Lemma B.1 on this subtree and reasoning as above, we have $t_u \leq \lceil \log_2 |V_u| \rceil \leq \lceil \log_2 |V| \rceil$.

As at least one factor node in the unicycle is peeled in round t_u, we must have that J_{t_u} is a tree or forest, which by Lemma B.1 can be peeled in at most $\lceil \log_2 |V| \rceil$ additional iterations, since the number of variable nodes in the J_{t_u} is at most $|V|$. Thus, $T_C(G) \leq t_u + \lceil \log_2 |V| \rceil$. Combining these two inequalities yields

$$T_C(G) \leq t_u + \lceil \log_2 |V| \rceil \leq 2\lceil \log_2 |V| \rceil. \qquad \square$$

PROOF OF LEMMA 5.5.   The lemma can be derived from known results (see, e.g., [7]), but we find it easier to provide an independent proof.

We use a generating function approach to prove the bound

$$(77) \qquad \mathbb{P}[Z_T > (\beta\theta)^T] \leq 2\exp(-C(\beta/2)^T).$$

Equation (35) follows (eventually for a different constant C) via union bound.

Define $f(s) \equiv \mathbb{E}[s^{Z_1}] = \sum_{j=0}^{\infty} s^j b_j$. By assumption, it is clear that $f(s)$ is finite for $s \in (0, 1/(1-\delta))$. Define $f^{(t)}(s) \equiv \mathbb{E}[s^{Z_t}]$ for $t \geq 1$ [so that $f(s) = f^{(1)}(s)$]. It is well known that

$$(78) \qquad f^{(t)}(s) = f(f^{(t-1)}(s))$$

for $\tau \geq 2$. It follows that $f^{(t)}(s)$ is finite for $s \in (0, 1/(1-\delta))$, and all $\tau \geq 2$.

By dominated convergence f is differentiable at 0 with $f'(0) = \theta$. Hence, there exists $\varepsilon_0 > 0$ such that, for all $\varepsilon \in [0, \varepsilon_0]$

$$(79) \qquad f(1+\varepsilon) \leq 1 + 2\theta\varepsilon.$$

By applying the recursion (78) and the fact that f is monotone increasing, we obtain, for all $\varepsilon \in [0, \varepsilon_0]$ obtain

$$(80) \qquad f^{(T)}(1+\varepsilon) \leq 1 + (2\theta)^T \varepsilon.$$

In particular setting $\varepsilon = \varepsilon_0/(2\theta)^T$, we get $f^{(T)}(1+\varepsilon) \leq 1 + \varepsilon_0 \leq 2$.

Finally, by Markov inequality,

$$\mathbb{P}\{Z_T \geq (\beta\theta)^T\} \leq (1+\varepsilon)^{-(\beta\theta)^T} f^{(T)}(1+\varepsilon)$$

$$\leq 2\left(1 - \frac{\varepsilon}{2}\right)^{(\beta\theta)^T} \leq 2e^{-(\beta\theta)^T/2},$$

which completes the proof. $\square$

APPENDIX C: PROOF OF TECHNICAL LEMMAS OF SECTION 6

PROOF OF LEMMA 6.1. We prove this lemma by induction. Let $B_{\mathrm{l}}^{(t)}$ and $B_{\mathrm{u}}^{(t)}$ be the result of t steps of backbone augmentation on graphs G_{s} and G with initial graphs $B_{\mathrm{l}}^{(0)}$ and $B_{\mathrm{u}}^{(0)}$, respectively. By assumption $B_{\mathrm{l}}^{(0)} \subseteq B_{\mathrm{u}}^{(0)}$. Now assume $B_{\mathrm{l}}^{(t)} \subseteq B_{\mathrm{u}}^{(t)}$. It is enough to show that if $a \in B_{\mathrm{l}}^{(t+1)} \setminus B_{\mathrm{l}}^{(t)}$ then $a \in B_{\mathrm{u}}^{(t+1)}$. Since $a \in B_{\mathrm{l}}^{(t+1)} \setminus B_{\mathrm{l}}^{(t)}$, we know that $a \in G$ and has at most one neighbor outside of $B_{\mathrm{l}}^{(t)}$. By induction assumption $B_{\mathrm{l}}^{(t)} \subseteq B_{\mathrm{u}}^{(t)}$ and, therefore, a has at most one neighbor outside $B_{\mathrm{l}}^{(t)}$. Hence, either $a \in B_{\mathrm{u}}^{(t)}$ or it is added to $B_{\mathrm{u}}^{(\infty)}$ at step $t+1$. □

PROOF OF LEMMA 6.6. Define $f(x) = 1 - \exp\{-k\alpha x^{k-1}\}$. It follows immediately from the definition of $\alpha_{\mathrm{d}}(k)$, that for $\alpha > \alpha_{\mathrm{d}}(k)$, we have $Q > 0$ and $f'(Q) \le 1$. Furthermore, a straightforward calculation yields

$$(81) \qquad f'(Q) = k(k-1)\alpha Q^{k-2} \exp\{-k\alpha Q^{k-1}\}.$$

It is therefore sufficient to exclude the case $f'(Q) = 1$. Solving the equations $f(Q) = Q$ and $f'(Q) = 1$, we get the following equation for Q:

$$(82) \qquad -(1-Q)\log(1-Q) = \frac{Q}{k-1},$$

which has a unique solution $Q_*(k)$ due to the concavity of the left-hand side. We can then solve for α yielding the unique value $\alpha = \alpha_*(k)$ such that $f(Q) = Q$ and $f'(Q) = 1$ admits a solution. On the other hand, these two equations are satisfied at $\alpha_{\mathrm{d}}(k)$ by a continuity argument. It follows that $\alpha_{\mathrm{d}}(k) = \alpha_*(k)$, and hence $f'(Q) < 1$ for all $\alpha > \alpha_{\mathrm{d}}(k)$. □

REFERENCES

[1] ACHLIOPTAS, D. and COJA-OGHLAN, A. (2008). Algorithmic barriers from phase transitions. In *Proc. of the 49th IEEE Symposium on Foundations of Computer Science, FOCS* 793–802. IEEE Computer Society, Los Alamitos, CA.

[2] ACHLIOPTAS, D., COJA-OGHLAN, A. and RICCI-TERSENGHI, F. (2011). On the solution-space geometry of random constraint satisfaction problems. *Random Structures Algorithms* **38** 251–268. MR2663730

[3] ACHLIOPTAS, D., NAOR, A. and PERES, Y. (2005). Rigorous location of phase transitions in hard optimization problems. *Nature* **435** 759–764.

[4] ACHLIOPTAS, D. and PERES, Y. (2004). The threshold for random k-SAT is $2^k \log 2 - O(k)$. *J. Amer. Math. Soc.* **17** 947–973 (electronic). MR2083472

[5] ALDOUS, D. and LYONS, R. (2007). Processes on unimodular random networks. *Electron. J. Probab.* **12** 1454–1508. MR2354165

[6] ALDOUS, D. and STEELE, J. M. (2004). The objective method: Probabilistic combinatorial optimization and local weak convergence. In *Probability on Discrete Structures* (H. Kesten, ed.). *Encyclopaedia Math. Sci.* **110** 1–72. Springer, Berlin. MR2023650

[7] ATHREYA, K. B. and NEY, P. E. (1972). *Branching Processes*. Springer, New York. MR0373040

[8] BALOGH, J., PERES, Y. and PETE, G. (2006). Bootstrap percolation on infinite trees and non-amenable groups. *Combin. Probab. Comput.* **15** 715–730. MR2248323

[9] BENJAMINI, I. and SCHRAMM, O. (1996). Percolation beyond $\mathbf{Z}_d$, many questions and a few answers. *Electron. Commun. Probab.* **1** 71–82 (electronic). MR1423907

[10] BOLLOBÁS, B. (1980). A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European J. Combin.* **1** 311–316. MR0595929

[11] BOLLOBÁS, B. (2001). *Random Graphs*, 2nd ed. *Cambridge Studies in Advanced Mathematics* **73**. Cambridge Univ. Press, Cambridge. MR1864966

[12] COCCO, S., DUBOIS, O., MANDLER, J. and MONASSON, R. (2003). Rigorous decimation-based construction of ground pure states for spin-glass models on random lattices. *Phys. Rev. Lett.* **90** 047205.

[13] COJA-OGHLAN, A. (2010). A better algorithm for random k-SAT. *SIAM J. Comput.* **39** 2823–2864. MR2645890

[14] DEMBO, A. and MONTANARI, A. (2008). Finite size scaling for the core of large random hypergraphs. *Ann. Appl. Probab.* **18** 1993–2040. MR2462557

[15] DEMBO, A. and MONTANARI, A. (2010). Gibbs measures and phase transitions on sparse random graphs. *Braz. J. Probab. Stat.* **24** 137–211. MR2643563

[16] DEMBO, A. and MONTANARI, A. (2010). Ising models on locally tree-like graphs. *Ann. Appl. Probab.* **20** 565–592. MR2650042

[17] DEMBO, A., MONTANARI, A. and SUN, N. (2013). Factor models on locally tree-like graphs. *Ann. Probab.* **41** 4162–4213. MR3161472

[18] DIETZFELBINGER, M., GOERDT, A., MITZENMACHER, M., MONTANARI, A., PAGH, R. and RINK, M. (2010). Tight thresholds for Cuckoo Hashing via XORSAT. In *Proc. of the 37th International Colloquium on Automata, Languages and Programming, ICALP. Lecture Notes in Computer Science* **6198** 213–225. Springer, Berlin.

[19] DUBHASHI, D. P. and PANCONESI, A. (2009). *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge Univ. Press, Cambridge. MR2547432

[20] DUBOIS, O. and MANDLER, J. (2002). The 3-XORSAT threshold. In *Proc. of the 43rd IEEE Symposium on Foundations of Computer Science, FOCS* 769–778. IEEE Computer Society, Los Alamitos, CA.

[21] FRIEDGUT, E. (1999). Sharp thresholds of graph properties, and the k-sat problem. *J. Amer. Math. Soc.* **12** 1017–1054. MR1678031

[22] HALL, P. (1982). *Rates of Convergence in the Central Limit Theorem. Research Notes in Mathematics* **62**. Pitman, London. MR0668197

[23] HOORY, S., LINIAL, N. and WIGDERSON, A. (2006). Expander graphs and their applications. *Bull. Amer. Math. Soc. (N.S.)* **43** 439–561 (electronic). MR2247919

[24] KALLENBERG, O. (2002). *Foundations of Modern Probability*, 2nd ed. Springer, New York. MR1876169

[25] KLEINJUNG, T., AOKI, K., FRANKE, J., LENSTRA, A. K., THOMÉ, E., BOS, J. W., GAUDRY, P., KRUPPA, A., MONTGOMERY, P. L., OSVIK, D. A., TE RIELE, H., TIMOFEEV, A. and ZIMMERMANN, P. (2010). Factorization of a 768-bit RSA modulus. In *Advances in Cryptology—CRYPTO* 2010. *Lecture Notes in Computer Science* **6223** 333–350. Springer, Berlin. MR2725602

[26] KRZAKALA, F., MONTANARI, A., RICCI-TERSENGHI, F., SEMERJIAN, G. and ZDE-BOROVÁ, L. (2007). Gibbs states and the set of solutions of random constraint satisfaction problems. *Proc. Natl. Acad. Sci. USA* **104** 10318–10323 (electronic). MR2317690

[27] LUBY, M., MITZENMACHER, M., SHOKROLLAHI, A. and SPIELMAN, D. A. (1998). Analysis of low density codes and improved designs using irregular graphs. In *Proc. of the* 30*th ACM Symposium on Theory of Computing*, STOC 249–258. ACM, New York.

[28] LUBY, M. G., MITZENMACHER, M., SHOKROLLAHI, M. A. and SPIELMAN, D. A. (2001). Efficient erasure correcting codes. *IEEE Trans. Inform. Theory* **47** 569–584. MR1820477

[29] MÉZARD, M. and MONTANARI, A. (2009). *Information*, *Physics*, *and Computation*. Oxford Univ. Press, Oxford. MR2518205

[30] MÉZARD, M., PARISI, G. and ZECCHINA, R. (2003). Analytic and algorithmic solution of random satisfiability problems. *Science* **297** 812–815.

[31] MÉZARD, M., RICCI-TERSENGHI, F. and ZECCHINA, R. (2003). Two solutions to diluted *p*-spin models and XORSAT problems. *J. Stat. Phys.* **111** 505–533. MR1972120

[32] MOLLOY, M. (2005). Cores in random hypergraphs and Boolean formulas. *Random Structures Algorithms* **27** 124–135. MR2150018

[33] MONASSON, R., ZECCHINA, R., KIRKPATRICK, S., SELMAN, B. and TROYANSKY, L. (1999). Determining computational complexity from characteristic "phase transitions". *Nature* **400** 133–137. MR1704845

[34] MONTANARI, A. (2013). Statistical mechanics and algorithms on sparse and random graphs. In *Lectures on Probability Theory and Statistics*. Saint-Flour.

[35] PITTEL, B., SPENCER, J. and WORMALD, N. (1996). Sudden emergence of a giant *k*-core in a random graph. *J. Combin. Theory Ser. B* **67** 111–151. MR1385386

[36] RICHARDSON, T. and URBANKE, R. (2008). *Modern Coding Theory*. Cambridge Univ. Press, Cambridge. MR2494807

[37] WORMALD, N. C. (1981). The asymptotic distribution of short cycles in random regular graphs. *J. Combin. Theory Ser. B* **31** 168–182. MR0630980

[38] WORMALD, N. C. (1999). Models of random regular graphs. In *Surveys in Combinatorics*, 1999 (*Canterbury*) (J. D. Lamb and D. A. Preece, eds.). *London Mathematical Society Lecture Note Series* **267** 239–298. Cambridge Univ. Press, Cambridge. MR1725006

M. IBRAHIMI
URBAN ENGINES
LOS ALTOS, CALIFORNIA 94022
USA
E-MAIL: ibrahimi@stanford.edu

M. KRANING
QADIUM, INC.
SAN FRANCISCO, CALIFORNIA 94107
USA
E-MAIL: matt@qadium.com

Y. KANORIA
DECISION, RISK AND OPERATIONS DIVISION
GRADUATE SCHOOL OF BUSINESS
COLUMBIA UNIVERSITY
NEW YORK, NEW YORK 10027
USA
E-MAIL: ykanoria@columbia.edu

A. MONTANARI
DEPARTMENT OF ELECTRICAL ENGINEERING
    AND DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305
USA
E-MAIL: montanari@stanford.edu