The solution space geometry of random linear equations

Dimitris Achlioptas
University of Athens & RACTI*

Michael Molloy
University of Toronto†

Abstract

We consider random systems of linear equations over GF(2) in which every equation binds k variables. We obtain a precise description of the clustering of solutions in such systems. In particular, we prove that with probability that tends to 1 as the number of variables, n, grows: for every pair of solutions σ, τ, either there exists a sequence of solutions $\sigma, \ldots, \tau$ in which successive elements differ by $O(\log n)$ variables, or every sequence of solutions $\sigma, \ldots, \tau$ contains a step requiring the simultaneous change of $\Omega(n)$ variables. Furthermore, we determine precisely which pairs of solutions are in each category. Our results are tight and highly quantitative in nature. Moreover, our proof highlights the role of unique extendability as the driving force behind the success of Low Density Parity Check codes and our techniques also apply to the problem of so-called pseudo-codewords in such codes.

*Research supported by an ERC IDEAS Starting Grant, an NSF CAREER Award, and a Sloan Fellowship.
†Dept. of Computer Science, University of Toronto. Research supported by an NSERC Discovery Grant.

1 Introduction

Random Constraint Satisfaction Problems (CSPs) have emerged as a mathematically tractable vehicle for studying the performance of algorithms and proof systems. In the most well-studied setting, one has a set of n variables all with the same (small) domain D and a set of $m = \Theta(n)$ constraints, each of which binds a randomly selected subset of $k = O(1)$ variables. Two canonical examples are random k-SAT and coloring sparse random graphs. Such random CSPs have proven difficult for proof systems, e.g., in the seminal work of Chvátal-Szeméredi on resolution [7], and, more recently, for some of the most sophisticated algorithms known [5, 6].

A fundamental quantity in the study of random CSPs is the so-called constraint density, i.e., the ratio of constraints-to-variables $\alpha = m/n$. In particular, for many NP-complete CSPs there is a critical density above which solutions provably exist, yet no known polynomial-time algorithm can find one. In [2] it was shown that for random k-SAT and for random graph coloring this algorithmic breakdown coincides with the onset of solution clustering: the set of solutions can be partitioned into exponentially many sets (clusters) that have linear Hamming distance from one another. Moreover, in each cluster a large fraction of the variables are frozen, i.e., take the same value in all solutions in the cluster. There is on-going work, see e.g., [21], to establish that clustering is a universal phenomenon for random CSPs.

Until now there has not been a precise description of the clusters for any random CSP. The main contribution of this paper is such a description of the clusters of k-XOR-SAT. This has long been recognized as one of the most accessible of the fundamental random CSP models, in that researchers have managed to prove difficult results for k-XOR-SAT that appear to be far beyond our reach for, e.g., random k-SAT and random graph coloring. Perhaps the most notable result along these lines is Dubois and Mandler's [9] determination of the satisfiability threshold for random k-XOR-SAT. Before we present our results we present some background on sparse systems of random linear equations over GF(2) that puts our results in perspective.

1.1 Random systems of linear equations

We consider systems of $m = O(n)$ linear equations over n Boolean variables, where each equation binds a constant number of variables $k \geq 3$. (The case $k \leq 2$ is trivial.) Clearly, deciding whether such a system has satisfying assignments (solutions) can be done in polynomial time by, say, Gaussian elimination. In fact, the set of solutions forms a subspace, so that the sum of two solutions is also a solution. At the same time, it seems that if one fails to exploit the underlying algebraic structure everything falls apart. For example, if the system is unsatisfiable, finding a value assignment σ that satisfies as many equations as possible, i.e., MAX XOR-SAT, is NP-complete. Moreover, given a satisfiable system and an arbitrary $\sigma \in \{0,1\}^n$, finding a solution nearest to σ is also NP-complete [3]. Finally, random k-XOR-SAT (defined below) appears to be extremely difficult both for generic CSP solvers and for SAT solvers working on a SAT encoding of the instance. Indeed, very recent work strongly suggests that among a wide array of random CSPs, random k-XOR-SAT is the *most* difficult for random walk type algorithms such as WalkSat [11].

Random k-XOR-SAT, which we study here, is the case where each equation binds exactly k variables. To form the random system of equations $Ax = b$ we take A to be the adjacency matrix of a random k-uniform hypergraph H with n variables and m edges and $b \in \{0,1\}^m$ to be a uniformly random vector. To choose A we can either select exactly m out of the possible $\binom{n}{k}$ edges uniformly and independently, or include each possible edge independently with probability p. (Results transfer readily betwen the two models when $m = p\binom{n}{k}$.) In this paper, we work with the latter model, which we denote $X_k(n,p)$. We will say that a sequence of events $\mathcal{E}_n$ holds *with high probability (w.h.p.)* for such a system if $\lim_{n\to\infty} \Pr[\mathcal{E}_n] = 1$.

We note that as $n \to \infty$, the degrees of the variables in such a random system tend to Poisson random variables with mean km/n. This means that for any finite $\alpha = m/n$, w.h.p. there will be $\Omega(n)$ variables of degree 0 and 1. Clearly, variables of degree 0 do not affect the satisfiability of the system. Similarly, if a variable v appears in exactly one equation e_i, then, we can always satisfy e_i by setting v appropriately for any constant b_i. Therefore, we can safely remove e_i from consideration and only revisit it after we have found a solution to the remaining equations. Crucially, this removal of e_i can cause the degree of other variables to drop to 1. This leads us to the definition of the core of a hypergraph.

Definition 1. *The* r-core *of a hypergraph* H *is the maximum subgraph of* H *in which every vertex has degree at least* r*.*

Trivially, removing any vertex of degree less than r from H does not change its r-core. Therefore, the r-core is the (potentially empty) outcome of the procedure: repeatedly remove an arbitrary vertex of degree less than r until no such vertices remain.

As discussed above, any satisfying assignment to the 2-core variables can be readily extended to the remaining variables. We will see in Section 2 below that there is a natural heuristic argument which would lead one to guess that different satisfying assignments on the 2-core variables have Hamming distance $\Omega(n)$. Indeed, in [16] rigorous but erroneous arguments were given for this. This guess motivates the following very simple definition of clusters: each cluster consists of all possible extensions to a given 2-core satisfying assignment. However, this guess is false, as we now describe:

Definition 2. *A* flippable cycle *in the 2-core of a hypergraph* H *is a set of vertices* $S = \{v_1, ..., v_t\}$ *in the 2-core with the following property: There is a set of edges* $e_1, ..., e_t$ *in the 2-core such that each vertex* v_i *lies in* e_i *and in* e_{i+1} *and in no other edges of the 2-core, for all* $1 \leq i \leq t$ *(addition mod* t*).*

Thus, the vertices $v_1, \ldots, v_t$ have degree two in the 2-core, whereas the remaining vertices in edges $e_1, ..., e_t$ can have arbitrary degree (at least 2) in the 2-core. Note that these remaining vertices are *not* part of the flippable cycle.

If σ is any 2-core satisfying assignment, then flipping the value of all variables in a flippable cycle readily yields another satisfying assignment of the 2-core. It is not hard to show that the 2-core of a random hypergraph often contains short flippable cycles, implying that 2-core satisfying assignments may have Hamming distance $\Theta(1)$, a far cry from the heuristic argument and the definition of clusters described above.

We note that the number of vertices in flippable cycles in the 2-core has mean $\Theta(1)$ (see Lemma 23).

In [16] it was also argued that for every pair σ, τ of extensions of the same 2-core assignment, there exists a sequence of satisfying assignments $\sigma, \sigma', \ldots, \tau$ such that successive assigments have Hamming distance at most $d = O(1)$. The argument in [16], though, has a grave flaw and in fact $O(\log n / \log\log n)$ is a lower bound on d (see Observation 4 below). This leaves open the question of how different 2-core extensions relate to one another (indeed, this was stated as an open problem in [17].) That is, is it possible to convert one into another by small (in Hamming distance) steps? Also, what about inside the 2-core? Is it possible to travel between different 2-core assignments by flipping a few variables at a time? In this work we give an *exact* answer to both of these questions, fully resolving the cluster structure of random k-XOR-SAT.

Definition 3. *Say that two solutions* σ, τ*:*

- *Are* cycle-equivalent *if on the 2-core they differ only on variables in flippable cycles in the 2-core (and arbitrarily on variables not in the 2-core).*
- *Are* d-connected *if there exists a sequence of solutions* $\sigma, \sigma', \ldots, \tau$ *such that the Hamming distance of every two successive elements in the sequence is at most* d*.*
- *Are* d-disconnected *if they are not* d*-connected.*

Let the **cluster** *of a solution* σ *consist of all its cycle-equivalent solutions.*

We first prove that in a sparse random system of linear equations, flippable cycles are w.h.p. the only source of connectivity between different 2-core assignments. That is, if two 2-core satisfying assignments differ on *even one* variable not lying in a flippable cycle, they must differ in $\Omega(n)$ variables.

Theorem 1. *For any constants* $c > 0$ *and* $k \geq 3$*, there exists a constant* $\alpha = \alpha(c, k) > 0$ *such that in* $X_k(n, p = c/n)$*, w.h.p. every pair of solutions in different clusters is* αn*-disconnected.*

In [16], Mézard et al. did prove that there exists a constant $\gamma > 0$ such that for any $\theta \in (0, \gamma)$, no two solutions differ on θn variables in the 2-core. However, they did not perform the crucial analysis for solutions that differ on $o(n)$ variables. This explains why they missed the fact that some solutions differ on the $o(n)$ vertices of a flippable cycle. The same erroneous statement appears in [17]; again the $o(n)$ analysis is missing. Providing that analysis, and using it to show that 2-core solutions which differ on $o(n)$ variables must differ *only* on flippable cycles, is the most difficult part of our proof of Theorem 1.

In stark contrast, we prove that internally clusters are very well-connected.

Theorem 2. *For any constants $c > 0$ and $k \geq 3$, there exists a constant $Q = Q(c, k) > 0$ such that in $X_k(n, p = c/n)$, w.h.p. every pair of solutions in the same cluster are $Q \log n$-connected.*

Our proof of Theorem 2 is algorithmic, giving an efficient method to travel between any pair of solutions in the same cluster. Theorem 2 is nearly tight due to the following.

Observation 4. *W.h.p. every cluster contains $g(n)$-disconnected solutions, where $g(n) = \Omega(\log n / \log \log n)$.*

Proof. Consider any solution σ to the 2-core, and consider any two extensions σ_0, σ_1 of σ to the entire formula such that, for a specific non-core variable v, we have $\sigma_0(v) = 0, \sigma_1(v) = 1$. Then σ_0, σ_1 must differ in at least one additional variable in every equation containing v. Thus their Hamming distance is at least $\deg(v) + 1$. The observation now follows readily since w.h.p. there are variables in the tree components of degree at least $g(n)$ and every variable in a tree-component can take both values in extensions of any σ. □

Finally, it is worth pointing out that for a large range of densities, the 2-core of a random k-uniform hypergraph is empty. For example, for $k = 3$ the 2-core is empty [23, 18] for $\alpha \leq 0.818...$, while unsatisfiability occurs [9] at $\alpha = 0.917...$ Since in the absence of a 2-core there is only one cluster, our Theorem 2 implies that in this regime the *entire* set of solutions is well-connected. To prove Theorem 2, we draw heavily from the linear structure of the constraints, (i) showing that it is possible to identify a set of non-core variables to act as a basis for expressing all possible extensions to each 2-core solution, and with a lot of work (ii) showing that changing any basis variable can only affect the value of $O(\log n)$ other variables.

So, in a nutshell, we prove that before the 2-core emerges any solution can be transformed to any other solution along a sequence of successive solutions differing in $O(\log n)$ variables. In contrast, as soon as the 2-core emerges, the set of solutions shatters into clusters defined by complete agreement on the 2-core variables not in flippable cycles: any two solutions that disagree on even one 2-core variable not in a flippable cycle, differ on $\Omega(n)$ variables. At the same time, solutions in the same cluster behave like solutions in the pre-core regime, i.e., one can travel arbitrarily inside each cluster by changing $O(\log n)$ variables at a time.

Our proof of Theorem 1 easily extends to all *uniquely extendable* CSPs.

Definition 5 ([8])**.** *A constraint of arity k is* uniquely extendable *if for any value assignment to any $k-1$ variables there is precisely one value for the unassigned variable that satisfies the constraint.*

Linear equations over GF(2) and unique games are the two most common examples of uniquely extendable (UE) CSPs, but many others exist (see, eg. [8]). Clearly, any instance of a UE CSP Φ is satisfiable iff its 2-core is satisfiable. Thus, it is natural to define clusters analogously to XOR-SAT, i.e., two solutions are in the same cluster if and only if their 2-core restrictions differ only on flippable cycles. Our proof of Theorem 1 applies readily to any UE CSP, yielding a corresponding theorem, i.e., that, under this definition of clusters, satisfying assignments in different clusters are $\Omega(n)$-disconnected (see the remark following Proposition 33). However, we do not know whether the analogue of Theorem 2 holds under this definition of clusters, i.e., whether it is possible to travel between cycle-equivalent solutions in small steps. Also, note that while in XOR-SAT changing all the variables in *any* flippable cycle results in another solution, this is not necessarily the case for every UE CSP Φ.

2 Earlier work, LDPC codes, and variables of degree 2

Consider a system of linear equations over GF(2), $Ax = b$. Note first that if the rows of A are linearly independent then the system has solutions for every choice of b, while if the rows are dependent, then it only has solutions for certain choices of b. It follows that the satisfiability threshold coincides with the threshold for the rows of A to be independent; if we are above that threshold, then it turns out that the rows are *highly* dependent and only a vanishingly small proportion of the choices for b will yield a solution.

One can use this observation to get an easy upper bound of $m/n \leq 1$ on the satisfiability threshold for random k-XOR-SAT by noting that if $m > n$ then the rows must be dependent. As one can imagine, though, this condition is not tight since variables of degree 0 and 1 only contribute fictitious degrees of freedom. The next most reasonable necessary condition for satisfiability then is $m_c \leq n_c$, where n_c, m_c is the number of variables and equations in the 2-core, respectively. In [9] Dubois and Mandler proved that this simple necessary condition is also sufficient by proving that the (exponential in expectation) number of solutions for the 2-core is strongly concentrated around its expectation.

Random sparse systems of linear equations are the backbone of modern coding theory as they underlie Low Density Parity Check codes (LDPC). To form such a code one first decides on its block length (the number of variables n), its rate (via the number of equations m), the fraction, λ_i, of variables of degree i, i.e., that will appear in i equations, and the fraction, ρ_i, of equations that will bind i variables (unlike k-XOR-SAT, different equations can bind different numbers of variables). Then, to get a concrete code, one chooses a uniformly random hypergraph satisfying these requirements using the configuration model (see Section 5), and sets its adjacency matrix to be A. The codewords then are the solutions to $Ax = 0$, since for any useful set of parameters, the equations are w.h.p. linearly independent, and so every choice of b is equivalent, making $b = 0$ a convenient choice.

Returning momentarily to random k-XOR-SAT we note that, by standard results, the 2-core of a random k-uniform hypergraph is itself a random k-uniform hypergraph conditional on its degree sequence. Moreover, w.h.p. the number of variables of each degree $i \geq 2$ in the 2-core is $\gamma_i \cdot n + o(n)$, where γ_i is a well-understood deterministic function [23, 18]. Therefore, we see that the 2-core of a random k-XOR-SAT instance is itself an LDPC code with $\lambda_i = \gamma_i$ and $\rho_i = k$ for all i.

To get an idea of how systems of random linear equations give rise to codes with good (typical) distance properties, it suffices to make two basic observations.

- Say that σ is a solution to a system as above, e is an equation in the system and v is a variable in e. If we flip v, in order to satisfy e we must now change (at least) one other variable u in e. Consider now any other equations containing u, v. In each of these, if we change u or v, at least one other variable must change. And so on. This propagation of forced changes stops only when all relevant loops close.

- If we insist that every variable appears in at least 3 equations, i.e., has degree at least 3, then the bipartite (factor) graph between variables and equations is an excellent expander. In particular, it is easy to prove that after changing a single variable the loops mentioned above close only after another $\Omega(n)$ variables have been changed (see, eg. [15]).

Clearly, in an error-correcting code there should be no variables of degree 0 or 1 since such variables do not carry *any* information about the message being sent (as any assignment to the remaining variables can be extended to them, they terminate any propagation sequence reaching them).

At the same time, it is natural to guess that if the minimum degree is 2, then the chains of propagations above will all be sufficiently large, motivating the (incorrect) definition of k-XOR-SAT clusters as the sets of solutions that agree on the 2-core. However, that guess is wrong. Degree 2 variables complicate matters by reducing the explosiveness of the propagation process, giving rise to a small number of pseudo-codewords (solutions near the $\mathbf{0}$ vector, ruining the code's worst-case distance). Moreover, we must tolerate this, because having a non-trivial fraction of degree 2 variables is crucial in making LDPC codes rate-efficient [17]. As we discuss in the next section, degree 2 variables also play a crucial and completely analogous role in the setting of this paper. Our flippable cycles give rise to pseudocodewords and Theorem 1 establishes that they are the only source of flaws in the distance properties of this code. To prove this theorem, we develop machinery

that can be applied to hypergraphs with *arbitrary* degree sequences of minimum degree 2. This yields a quantitative analysis of pseudocodewords which we will present in another paper.

3 Proof Outline

3.1 Theorem 1: Cluster separation

Given a satisfying assignment σ, a *flippable set* is a set of variables S such that flipping the value of all variables in S yields another satisfying assignment τ. Proving Theorem 1 boils down to proving that w.h.p. every flippable set in the 2-core other than a flippable cycle has linear size.

A common approach to proving analogous statements is to establish, deterministically, that every flippable set must induce a dense subgraph. In particular, if one can prove that for some constant $\epsilon > 0$, every such set is at least $1+\epsilon$ times as dense as a flippable cycle, then standard arguments yield the desired conclusion. Here, though, this is not the case, due to the possibility of arbitrarily long 2-linked paths (see Definition 31 below). Specifically, it is easy to see that by replacing every edge of a flippable set by a long 2-linked path, one can create a new flippable set whose density is arbitrarily close to that of a flippable cycle. Thus, controlling the number and interactions of 2-linked paths, an approach similar to that of [1, 20], is crucial to our argument. Carrying out this analysis on hypergraphs with a given degree sequence, as we do here, requires the introduction of a number of technical innovations.

The key to controlling 2-linked paths is to bound a parameter governing the degree to which they tend to branch. Lemma 20 shows that this parameter is bounded below 1, so while arbitrarily long 2-linked paths will occur, their frequency decreases exponentially with their length. This enables us to tame them.

We note that if we were working on hypergraphs with minimum degree at least 3, then there would be no 2-linked paths, and the proof would have been very easy. All of our innovations were designed to deal with the problem of degree 2 vertices and our approach readily applies to arbitrary degree sequences of minimum degree 2. In particular, since our analysis is developed for arbitrary degree sequences it provides a new tool to attack the issue of pseudocodewords in LDPC codes, which we do in a separate work. This is an important point since, as mentioned above, having a linear number of vertices of degree 2 is an essential requirement for any such code to approach channel capacity.

3.2 Theorem 2: Connectivity inside clusters

Recall that we can reach the r-core of a hypergraph by repeatedly removing any one vertex of degree less than r, until no such vertices remain.

Definition 6. *An* r-stripping sequence *is a sequence of vertices that can be deleted from a hypergraph, one-at-a-time, such that at the time of deletion, each vertex has degree less than* r*.*

It is often useful to consider stripping the vertices in several rounds.

Definition 7. *The* parallel r-stripping process *consists of iteratively removing* all *vertices of degree less than* r *at once along with any hyperedges containing any of those vertices, until no vertices of degree less than* r *remain.*

It is easy to show using standard facts about r-cores of random hypergraphs that for every constant $\epsilon > 0$, there is a constant $T = T(\epsilon)$ such that in a random k-uniform hypergraph, all but ϵn of the non-core vertices can be removed by a stripping sequence of length at most T. What is significantly harder, and our main technical contribution in order to establish Theorem 2, is proving that w.h.p. *all* non-core vertices can be removed by a stripping sequence of length $O(\log n)$. (See Section 4 for a description of why this suffices to prove Theorem 2.) We note that this result is of independent interest in random hypergraph theory and for that reason we prove it for arbitrary $r \geq 2$, even though we only need the case $r = 2$ case for Theorem 2.

To prove that all non-core vertices can be removed by a stripping sequence of length $O(\log n)$, our approach is significantly different below and above the threshold, $c^*_{k,r}$ for the emergence of an r-core in

random k-uniform hypergraphs. In both cases, we begin by stripping down to H_B for a sufficiently large constant B. A simple argument shows that for any non-core vertex v, the number of vertices removed during this initial phase that are relevant to the removal of v, is bounded. What remains is to show that every non-core vertex in H_B can be removed *from* H_B by a stripping sequence of length $O(\log n)$.

For $c < c^*_{k,r}$, we prove that there exists a sufficiently large constant $B = B(c,k,r)$ such that after B rounds of the parallel r-stripping process, all connected components of the remaining hypergraph H_B have size at most $W = O(\log n)$; therefore, all remaining vertices can be removed with an additional W strips. To do this we establish analytic expansions for the degree sequence of H_B as B grows and then apply a hypergraph extension of the main result of Molloy and Reed [19] regarding the component sizes of a random k-uniform hypergraph with a given degree sequence.

For $c > c^*_{k,r}$, a lot more work is required. Once again, 2-linked paths are a major problem. Indeed, it is not hard to see that a long 2-linked path with one endpoint of degree 1, can create a long stripping sequence leading to the removal of its other endpoint.

We first establish that for any $\epsilon > 0$, after a finite number $B = B(\epsilon)$ of r-stripping rounds we are sufficiently close to the r-core for two important properties to hold in the remaining hypergraph H_B: (i) there are at most ϵn vertices of degree less than r, and (ii) the "branching" parameter for 2-linked paths, mentioned above, is bounded below 1. Property (ii) allows us to control long 2-linked paths. However, this does not suffice as we need to control, more generally, for large tree-like stripping sequences. To do so, we note that any large tree must either have many leaves, or long paths of degree 2 vertices. Such long paths will correspond to 2-linked paths in the random hypergraphs, and so (ii) allows us to control the latter case. Leaves of the tree will have degree less than r, and so (i) enables us to control the former case.

4 Selecting a basis for intra-cluster travel

Given an r-stripping sequence that removes every vertex outside of the r-core, we associate with it a directed graph as follows: (i) let v be the next vertex in the sequence; (ii) remove the (no more than $r-1$) edges E_v containing v; (iii) add arcs to v from every vertex other than v in E_v. Note that if E_v contains a variable $w \neq v$ of current degree 1, w will end up having in-degree 0, as the edges in E_v are erased by the processing of v. More generally, no arcs are associated with the removals of degree 0 variables. So the digraph will include all vertices that are not in the r-core, as well as any r-core vertex that shares an edge with a vertex not in the r-core.

For every vertex v in the digraph, we define $R^+(v)$ to be the set of vertices that can be reached from v. The following lemma is the heart of our proof of Theorem 2 and its proof occupies Sections 7 and 8, after we set out some basic facts about cores in Section 5 and some basic calculations in Section 6.

Lemma 8. *Fix integers k, r and let $H = H_k(n,p)$ be a random k-uniform hypergraph on n vertices with $p = c/n^{k-1}$. There exists a constant $Q > 0$ such that w.h.p. there is an r-stripping sequence of H culminating with the r-core such that in the digraph D associated with the sequence:*

(a) For every vertex v, $|R^+(v)| \leq Q \log n$.

(b) For $r = 2$, for every flippable cycle C,

$$\sum_{v \in C} |R^+(v)| \leq Q \log n \ .$$

We note that, via standard arguments, a number of earlier works implicitly establish that for any fixed $\theta > 0$, there exists $Z = Z(\theta)$ such that $|R^+(v)| \leq Z$ for all but θn vertices v. The difficulty here lies in proving a bound on $|R^+(v)|$ that holds for **all** vertices.

Remark: For our purposes, we only need Lemma 8 for the case $r = 2, k \geq 3$. However, this lemma is of independent interest to the study of random hypergraphs, and so we prove it for all $r, k \geq 2$.

Remark: The proof of Lemma 8 can be extended to show that w.h.p. for every vertex $v \in D$, the subgraph induced by $|R^+(v)|$ has at most as many arcs as vertices.

Proof of Theorem 2. Given an arbitrary system of linear equations consider any 2-stripping sequence $v_1, ..., v_t$, of its associated hypergraph which strips all the way to the 2-core. Let D be the digraph associated with the sequence. Let $e_1, e_2, \ldots, e_t$ be the corresponding sequence of removed hyperedges, i.e. e_i was the only edge containing v_i when v_i was removed. Let B be the set of non-core vertices with indegree zero in D.

We will perform Gaussian elimination in an manner such that for any satisfying 2-core assignment, σ, B is a basis for the subspace formed by all satisfying extensions of σ. Specifically, we consider the removed hyperedges in reverse removal order, i.e., $e_t, e_{t-1}, \ldots, e_1$. Each time we consider a hyperedge e_i, we express v_i in terms of the other variables in e_i. Each of these other variables is either in B or in the 2-core, or is already expressed in terms of variables in B and the 2-core. So, in the end, each non-core variable $u \notin B$ is expressed in the form $u = \sum_{w \in \chi(u)} w$, where $\chi(u)$ is a set containing variables from the 2-core and B that have a path to u in D.

More specifically, for each $v \in B$ and for each v in the 2-core, we set $\chi(v) = \{v\}$. Every non-core variable not in B has indegree at least 1 and so is v_i for some removed hyperedge e_i. For each other $u \in e_i$, either $u \in B$, or u is in the 2-core, or $u = v_j$ for some $j > i$. When processing the edges in reverse removal order, i.e. $e_t, \ldots, e_1$, to process e_i we set $\chi(v_i)$ to the disjoint union of the sets $\chi(u)$, over all $u \in e_i$ other than v_i. That is, a variable $z \in \chi(v_i)$ iff $z \in \chi(u)$ for an odd number of variables $u \in e_i$ other than v_i. Since e_i is the equation $v_i = \sum_{u \in e_i; u \neq v} u$, by induction, $v_i = \sum_{w \in \chi(v_i)} w$. Note also that, by induction, $\chi(v_i)$ contains only vertices that are in B or the 2-core. Finally, note that possibly $\chi(v_i) = \emptyset$; in that case, $v_i = \sum_{w \in \chi(v_i)} w = 0$ in every satisfying assignment.

Remark: It is not hard to adapt the proof of Lemma 8 to show that w.h.p. for every i, $\chi(v_i) \neq \emptyset$. But that is not required for the purposes of this paper.

If at this point we fix any satisfying assignment σ to the 2-core variables, the variables in B will remain unrestricted. Moreover, since all non-core variables not in B are expressed in terms of some subset of variables in the union of the 2-core and B, the variables in B form a basis for the subspace corresponding to all possible extensions of σ.

Since B is a basis, there are exactly $2^{|B|}$ extensions of σ to the non-core variables, one for each assignment to B. We can move between any two such extensions by changing the assignments to the variables of B, one-at-a-time. Each time we change a variable $v \in B$, in order to get to another satisfying assignment, we only need to change a subset of $R^+(v)$ in the digraph D, because only variables $u \in R^+(v)$ can have $v \in \chi(u)$. Thus, by Lemma 8(a), we can move between any two such solutions changing at most $Q \log n$ variables at a time.

To complete the proof of Theorem 2, we show that we can move between any two cycle-equivalent solutions to the 2-core by changing a small number of variables. We move between any two such solutions by switching one flippable cycle at a time. Switching a flippable cycle, S, may require also changing some non-core variables in $\cup_{u \in S} R^+(u)$. By Lemma 8(b), this requires switching at most $Q \log n$ variables. □

4.1 Frozen variables

We close this section by showing how the digraph D can also be used to determine all of the frozen variables. Recall that a variable is said to be frozen in a cluster, if it takes the same value in all assignments of the cluster. In general, e.g., in random k-SAT, the set of frozen variables can differ from cluster to cluster. In random k-XOR-SAT, though, the set of frozen variables depends only on the underlying hypergraph, i.e., is the same for all clusters. Specifically, the 2-core variables not in flippable cycles are frozen by definition, whereas all 2-core variables in flippable cycles are not frozen. Below, we determine which non-core variables are also frozen and, we will see, the answer is independent of the 2-core assignment.

To obtain a basis for all core-solutions, prior to seeking a basis for the extensions we form a basis for the possible assignments to the flippable cycles. To do this, we first choose, for each flippable cycle C, one variable v_C arbitrarily and use $|C| - 1$ edges of the cycle to eliminate all variables in C other than v_C. Thus, for every variable $v \in C$ other than v_C, we have $v = v_C + z$, where $z \in \{0, 1\}$ depends on the cluster; we set $\chi(v) = \{v_C\}$. Clearly, for any cluster, the set of chosen variables forms a basis S for the subspace of its core solutions. Next, we apply the method, from the proof of Theorem 2, for finding a basis for the set of

all extensions. The result is a set B containing S and non-core variables.

Theorem 3. *In every cluster, the frozen variables consist of the 2-core variables not in flippable cycles, and the non-core variables whose set χ does not contain any variable from B.*

Proof. It suffices to prove that B is a basis for all solutions in the cluster. This follows trivially from the fact that S is a basis for the core solutions of the cluster, and the fact that our procedure in the proof of Theorem 2 yields a basis for all possible extensions of any 2-core solution. □

5 Random hypergraphs and their cores

We will use the configuration model of Bollobás [4] to generate a random k-uniform hypergraph H with a given degree sequence. Suppose we are given the degree $d(v)$ for each vertex v; thus $\sum d(v) = kE$ where E is the number of hyperedges. We take $d(v)$ *copies* of each v, and we take a uniformly random partition of these kE vertex-copies into E sets of size k. This naturally yields a k-uniform hypergraph, by mapping each k-set to a hyperedge on the vertices whose copies are in the k-set. Note that the hypergraph may contain loops (two copies of the same vertex in one hyperedge) and multiple edges (two identical hyperedges). It is well-known that the probability that this partition yields a simple hypergraph (i.e., one with no loops or multiple edges) is bounded below by a constant for degree sequences[1] satisfying certain conditions:

Definition 9. *Say that a degree sequence $\mathcal{S}$ is* nice *if $E = \Theta(n)$ and $\sum_v d(v)(d(v)-1) = O(n)$.*

Every degree sequence we will consider will correspond to some subgraph of $H_k(n,p)$ with a linear expected number of edges. Since, as is well-known, the degree sequence of such random hypergraphs is nice w.h.p., all the degree sequences we will consider will be nice. With this in mind, we will make heavy use of the following standard proposition and corollary as working in the configuration model is technically much easier than working with uniformly random hypergraphs with a given degree sequence.

Proposition 10. *If $\mathcal{S}$ is a nice degree sequence, then there exists $\epsilon > 0$ such that the probability that a random hypergraph with degree sequence $\mathcal{S}$ in the configuration model is simple is at least ϵ.*

This immediately yields:

Corollary 11. *If $\mathcal{S}$ is a nice degree sequence then:*

- *If property Q holds w.h.p. for k-uniform hypergraphs with degree sequence $\mathcal{S}$ in the configuration model, then Q holds w.h.p. for uniformly random simple hypergraphs with degree sequence $\mathcal{S}$.*
- *For any random variable X, if $E(X) = O(1)$ for k-uniform hypergraphs with degree sequence $\mathcal{S}$ in the configuration model, then $E(X) = O(1)$ for uniformly random simple hypergraphs with degree sequence $\mathcal{S}$.*

The following lemma will be very useful. Its exponential term is not tight, but will suffice for our purposes.

Lemma 12. *Consider a random k-uniform configuration with E edges, i.e., with total degree kE. For each $i = 2, ..., k$, specify ℓ_i sets of i vertex-copies, and set $L = \sum_{i=2}^{k} \ell_i$. The probability that each of these sets appears in some hyperedge, and no two appear in the same hyperedge is less than*

$$\exp\left(\frac{kL^2}{E-L}\right)\prod_{i=2}^{k}\left(\frac{(k-1)(k-2)...(k-i+1)}{(kE)^{i-1}}\right)^{\ell_i} .$$

[1] Clearly, we are referring to a sequence of degree sequences $\mathcal{S}_n$ so that asymptotic statements are meaningful. We suppress this point though, throughout, to streamline exposition.

Proof. We choose the hyperedges of the configuration by processing the specified sets one-at-a-time. To process one of the ℓ_i sets of size i, we first choose one set member γ arbitrarily and then randomly select the remaining $k-1$ vertex-copies of the hyperedge containing γ. Every time we do this there are at least $kE-kL$ yet unselected vertex-copies. Thus, the probability we chose all other $i-1$ members of the specified set is at most

$$\begin{aligned}\frac{(k-1)(k-2)...(k-i+1)}{(kE-kL)^{i-1}} &< \frac{(k-1)(k-2)...(k-i+1)}{(kE)^{i-1}} \times \left(\frac{E}{E-L}\right)^{i-1} \\ &< \frac{(k-1)(k-2)...(k-i+1)}{(kE)^{i-1}} \times e^{kL/(E-L)},\end{aligned}$$

since $i \le k$. So the probability that all L tuples are chosen to be in a hyperedge is less than

$$\prod_{i=2}^{k}\left(\frac{(k-1)(k-2)...(k-i+1)}{(kE)^{i-1}}\right)^{\ell_i} \times e^{(kL/(E-L))\ell_i} = e^{kL^2/(E-L)} \times \prod_{i=2}^{k}\left(\frac{(k-1)(k-2)...(k-i+1)}{(kE)^{i-1}}\right)^{\ell_i}$$

□

5.1 Cores

It is well-known that the r-core of a random k-uniform hypergraph is uniformly random conditional on its degree sequence. See [22] for the case $k=2$, and [18] for the nearly identical proof for general r. In fact, the same is true of the graph remaining after any number of iterations of the parallel stripping process. It is also straightforward (see e.g., [18]) to show the following propositions.

Proposition 13. *Let $H = H_k(n,p)$ be a random k-uniform hypergraph and let $H = H_0, H_1, \ldots$ be the sequence of hypergraphs produced by the parallel stripping process.*

(a) For every $i \ge 0$, H_i is uniformly random with respect to its degree sequence.

(b) There exist functions $\rho_0, \rho_1, \ldots$ such that for any fixed integer s, w.h.p. H_s contains $\rho_j(s)n + o(n)$ vertices of degree j and $\frac{1}{k}(\sum_{j\ge 1} j\rho_j(s))n + o(n)$ edges.

The following is a bit stronger than Proposition 13(a).

Proposition 14. *Let $H = H_k(n,p)$ and for any $i \ge 0$, let H_i be the hypergraph produced by i rounds of the parallel stripping process. Expose $V(H_i)$, and for each $0 \le j \le r-1$, expose V_j, the set of vertices of degree j in H_i. Thus $V_{\ge r} = V(H_i) - \cup_{j=0}^{r-1} V_i$ is the set of vertices of degree at least r in H_i. Finally, expose E, the number of edges in H_i. The hypergraph H_i is uniformly random conditional on these parameters.*

Proof. Consider any two hypergraphs A, A' on $V(H_i)$ which satisfy these parameters. Consider any hypergraph H such that, if we apply the parallel stripping process for i rounds, then $H_i = A$. Create H' by replacing A, the graph induced by $V(H_i)$, with A'. Since A, A' agree on $V_0, ..., V_{r-1}$, it is easy to see that applying the parallel stripping process to H' for i rounds will leave $H_i' = A'$. Furthermore, H, H' are both equally likely to be chosen as $H_k(n,p)$, since they have the same number of edges. Thus A, A' are equally likely to be H_i. □

Proposition 13 allows us to use the configuration model to study H_i. We will begin by bounding the contribution of the highest degree vertices to the sum of the degrees.

Lemma 15. *For every $\theta > 0$ there exists $J = J(\theta)$ such that for all i,*

$$\left|\sum_{v\in H_i} \deg_{H_i}(v) - \sum_{j=1}^{J} j\rho_j(i)\right| < \theta n \ .$$

Proof. By Proposition 13, the difference is $o(n) + \sum_{v:\deg_{H_i}(v)>J} \deg_{H_i}(v)$. Since $H_i \subseteq H_0 = H$, this sum is bounded by the corresponding sum in H, i.e., $\sum_{v:\deg_H(v)>J} \deg_H(v)$. The fact that this sum can be made arbitrarily smaller than θn for any $\theta > 0$ by taking $J = J(\theta)$ sufficiently large, is well-known and follows from the facts that (i) for each constant i, the number of vertices of degree i in H is w.h.p. $\lambda_i n + o(n)$ for a particular $\lambda_i = \lambda_i(c,k)$ and (ii) the number of hyperedges in H is highly concentrated around $(\frac{1}{k}\sum_{i\geq 1} i\lambda_i)n$. □

The following similar bound will also be useful:

Lemma 16. *For every constant d and $\theta > 0$ there exists $J = J(d,\theta)$ such that for all i,*

$$\left| \sum_{v:\deg_{H_i}(v)\geq d} \frac{\deg_{H_i}(v)!}{(\deg_{H_i}(v)-d)!} - \sum_{j=d}^{J} \frac{j!}{(j-d)!}\rho_j(i) \right| < \theta n \ .$$

Proof. The proof is almost identical to that of Lemma 15 but exploits the concentration of the number of d-stars in H, rather than of the number of edges. (Recall that a d-star is a set of d edges which contain a common vertex.) The concentration of the number of d-stars in H is easily established, e.g., by the Second Moment Method or Talagrand's Inequality. (Indeed, Lemma 15 and its proof are special cases of this lemma and its proof for $d = 1$.) □

For any fixed integers k, r and real number $\lambda > 0$, we write

$$\Psi_r(\lambda) = e^{-\lambda} \sum_{i\geq r-1} \lambda^i/i! \qquad \text{and} \qquad f_{k,r}(\lambda) = f(\lambda) = \frac{(k-1)!\lambda}{\Psi_r(\lambda)^{k-1}} \ .$$

See [18, 14, 10, 12] for proofs that the threshold for the appearance of an r-core in a random k-uniform hypergraph $H_k(n,p)$ with $p = c/n^{k-1}$ is

$$c^*_{k,r} = \min_{\lambda>0} f_{k,r}(\lambda).$$

We will see that f' has a unique 0 and, thus, for $c > c^*_{k,r}$ the equation $f(\lambda) = c$ has two solutions.

Definition 17. *For $c > c^*_{k,r}$, let $\mu = \mu(c)$ denote the larger of the two solutions of $f(\lambda) = c$.*

The following two propositions are well-known; see eg. [18] for proofs.

Proposition 18. *For every fixed $j \geq r$, w.h.p. the r-core contains $(e^{-\mu}\mu^j/j!)n + o(n)$ vertices of degree j. Furthermore, w.h.p. the r-core contains $(\mu/k)\Psi_r(\mu)n + o(n)$ edges.*

Proposition 19. *For every $\theta > 0$, there exists $B = B(\theta)$ such that w.h.p.*

(a) H_B contains fewer than θn vertices not in the r-core;

(b) For each $j \geq r$, $|\rho_j(B) - e^{-\mu}\mu^j/j!| < \theta$.

The following lemma will be critical for our analysis.

Lemma 20. *For every $c > c^*_{k,r}$, there exists $\zeta = \zeta(k,r,c) > 0$ such that*

$$(k-1)\frac{\mu^{r-1}}{(r-2)!} < (1-\zeta)\sum_{i\geq r-1} \frac{\mu^i}{i!} \ , \tag{1}$$

where μ is the largest of the two roots of the equation $f_{k,r}(\lambda) = c$.

Proof.

$$f'(\lambda) = 0 \iff \Psi_r(\lambda) = \lambda(k-1)\Psi_r(\lambda)^{k-2}\Psi'_r(\lambda) \iff \sum_{i\geq r-1} \frac{\lambda^i}{i!} = (k-1)\frac{\lambda^{r-1}}{(r-2)!} \ . \tag{2}$$

Equation (2) yields $c^*_{k,r} = f(\lambda^*)$ for some λ^* satisfying the last equation in (2). For $c > c^*_{k,r}$, since $\mu = \mu(c)$ is the largest of the two roots of $f(\lambda) = c$, it follows that $\mu > \lambda^*$. The lemma now follows by noting that the RHS of (1) divided by the LHS is proportional to $\sum_{i\geq r-1} \frac{\mu^{i-r+1}}{i!}$, which is clearly increasing with μ. □

6 Preliminaries to the proof of Lemma 8

As we said above, we will choose a sufficiently large constant B, strip down to H_B, and then focus on $R^+(u) \cap H_B$, making use of the fact that H_B is very close to the 2-core (by Proposition 19). The following will be used to bound the number of vertices that are removed from $R^+(u)$ when stripping down to H_B. For integer $s \geq 0$, we use $N^s(v)$ to denote the s-th neighborhood of v, i.e., the set of vertices within distance s from v. For any set of vertices A, $N^s(A) = \bigcup_{v \in A} N^s(v)$. We consider a single vertex to be a connected set. A straightforward induction yields the following.

Proposition 21. *For any integer i and vertex $u \in H_i$, $R^+(u) \subseteq N^i(R^+(u) \cap H_i)$.*

Lemma 22. *For any $c, s \geq 0$, there exists $\Gamma = \Gamma(c,s)$ such that in a random graph $G(n,p)$ with $p = c/n$, w.h.p. for every connected subset A of vertices $|N^s(A)| \leq \Gamma(|A| + \log n)$.*

Proof. We prove this for the case $s = 1$, i.e., that there is a constant $\gamma > 1$ such that w.h.p. every connected subset of vertices A satisfies $|N(A)| \leq \gamma(|A| + \log n)$. By iterating, we obtain that for every $s \geq 1$, every connected subset of vertices A satisfies $|N^s(A)| \leq f_s(|A|)$ where

$$\begin{aligned} f_1(x) &= \gamma(x + \log n) \\ f_{i+1}(x) &= \gamma(f_i(x) + \log n), \text{ for } i \geq 1. \end{aligned}$$

A simple induction yields $f_i(x) \leq \gamma^i(x + i \log n)$ and that yields the lemma with $\Gamma = s\gamma^s$.

Given any set A of size a, the probability that A is connected is at most the expected number of spanning trees which is $a^{a-2}(c/n)^{a-1}$. After conditioning that A is connected, the number of neighbors outside of A is distributed as $\mathrm{Bin}(a(n-a), c/n)$. The probability that this exceeds z is at most

$$\binom{a(n-a)}{z} \left(\frac{c}{n}\right)^z < \left(\frac{eca}{z}\right)^z < 2^{-z}, \qquad \text{for } z > 2eca.$$

For any $\gamma > 2$, if $|N(A)| > \gamma(|A| + \log n)$, then we must have $|N(A) \backslash A| > \frac{1}{2}\gamma(|A| + \log n)$. Taking $\gamma > 4ec$, the expected number of connected sets A satisfying this last inequality is at most

$$\binom{n}{a} a^{a-2} \left(\frac{c}{n}\right)^{a-1} 2^{-\frac{1}{2}\gamma(a+\log n)} < \frac{en}{a^2} (ec)^{a-1} 2^{-\frac{1}{2}\gamma(a+\log n)} < \frac{en}{a^2} \left(\frac{ec}{2^{\gamma/2}}\right)^{a-1} 2^{-\frac{1}{2}\gamma \log n} = n^{-\Theta(\gamma)} \ ,$$

for γ sufficiently large. Multiplying by the n choices for a yields the lemma. □

Lemma 23. *Fix $k \geq 3$ and let $H = H_k(n,p)$ be a random k-uniform hypergraph with $p = c/n^{k-1}$, where $c > c^*_{k,2}$. The expected number of vertices in flippable cycles in the 2-core of H is $O(1)$.*

Proof. Let $\mathcal{D}$ be the degree sequence of the 2-core of H. By Corollary 11, we can work in the configuration model. Recalling Definition 17, Proposition 18 and Lemma 20, w.h.p.

(i) $\mathcal{D}$ has total degree $\gamma n + o(n)$, where $\gamma = \mu \Psi_r(\mu)$,

(ii) $\mathcal{D}$ has $\lambda_2 n + o(n)$ vertices of degree 2, where $\lambda_2 = e^{-\mu}\mu^2/2$,

(iii) there exists $\zeta > 0$ such that $2(k-1)\lambda_2 < (1-\zeta)\gamma$.

We first bound the expected number of flippable cycles of size a in the 2-core. Let $\Lambda = \gamma n + o(n)$ be the total number of vertex copies, and let $L = \lambda_2 n + o(n)$ be the number of copies of degree 2 vertices.

There are $\binom{L}{a}$ choices for the connecting vertices, $\frac{(a-1)!}{2}$ ways to order them into a cycle, and 2^a ways to align their vertex-copies. This yields a pairs $\{y_1, z_1\}, \ldots \{y_a, z_a\}$ of vertex copies, each of which must land in a hyperedge. We process these pairs one-at-a-time, halting if we ever find that the pair does not land in a hyperedge. To process pair i, we ask only whether z_i lands in the same hyperedge as y_i; if it does we do *not* expose the other vertex-copies in that hyperedge. Thus, prior to processing pair i, we have exposed exactly

$2i-2$ vertex-copies, all of degree 2. There are $k-1$ other copies appearing in the same hyperedge as y_i. Each of the $\Lambda-(2i-1)$ unexposed copies (not including y_i) is equally likely to be one of those copies (and, for $k \geq 3$, the exposed copies also have positive probability). So the probability that z_i is one of them is at most $(k-1)/(\Lambda-2i+1)$. So the expected number of flippable cycles of length a is at most:

$$\binom{L}{a}\frac{(a-1)!}{2}2^a\prod_{i=1}^{a}\frac{k-1}{\Lambda-2i+1} < \tfrac{1}{2a}\prod_{i=1}^{a}\frac{2(k-1)(L-i+1)}{\Lambda-2i+1}.$$

By condition (iii) above, $2(k-1)L/(\Lambda-1) < 1-\frac{1}{2}\zeta$, and so $2(k-1)(L-i+1)/(\Lambda-2i+1) < 1-\frac{1}{2}\zeta$ for each i, since $L \leq \frac{1}{2}(\Lambda-1)$. So the expected number is at most $\frac{1}{2a}(1-\frac{1}{2}\zeta)^a$, and so the expected total number of vertices on flippable cycles is at most $\frac{1}{2}\sum_{a\geq 1}(1-\frac{1}{2}\zeta)^a = O(1)$. □

7 Proof of Lemma 8 above the r-core threshold

Let $H = H_k(n,p)$ be a random k-uniform hypergraph with $p = c/n^{k-1}$ and let $H = H_0, H_1, \ldots$ be the sequence of hypergraphs produced by the parallel r-stripping process. We will choose a stripping sequence that is consistent with the parallel process; i.e., in our stripping sequence: for every $i < j$, the vertices deleted in round i of the parallel process come before the vertices deleted in round j of the parallel process.

Let D be the digraph associated with this r-stripping sequence for H and recall that $R^+(u)$ denotes the set of vertices reachable from a vertex u in D.

7.1 Proof of part (a)

Our main challenge is to prove the following lemma. The idea is that we will take B large enough so that by stripping down to H_B, Proposition 19 gives us control of the degree sequence that remains, and Lemma 20 allows us to prove that a certain branching process involving long paths in a graph constructed from H_B dies out.

Lemma 24. *For every $c > c^*_{k,r}$ there exists $B = B(c)$ and $Q = Q(B,c)$ such that w.h.p. for every vertex u, $|R^+(u) \cap H_B| \leq Q\log n$.*

Proof of Lemma 8(a). Consider any vertex u. If $u \notin H_B$, then by Proposition 21, $R^+(u) \subseteq N^B(u)$ in which case Lemma 22 immediately implies that $|R^+(u)| < \Gamma(1+\log n)$ for some constant $\Gamma = \Gamma(c,B)$.

If $u \in H_B$, then $R^+(u) \subseteq N^B(R^+(u) \cap H_B)$, by Proposition 21. Since, by Lemma 24, $|R^+(u) \cap H_B| \leq Q\log n$, Lemma 22 now implies that $|R^+(u)| < \Gamma(Q\log n + \log n) = Z\log n$ for $Z = \Gamma Q + 1 = Z(c,B)$. □

For any i, we define D_i to be the subdigraph of D induced by the vertices in H_i. We also define:

Definition 25. *For any vertex $u \in D_i$, let $S_i(u)$ denote $R^+(u) \cap H_i$.*

Let T be a BFS tree in D_i rooted at u, thus spanning the vertices of $S_i(u)$. Since T is a BFS tree, each vertex has indegree at most 1 in T implying that no two arcs of T were formed during the removal of the same hyperedge. From this point on, we will treat T as an undirected graph.

The following technical lemma bounds the density of small subgraphs of $H_k(n,p)$. Lemmas of this flavour are very common in random graph theory. Given a subset S of the vertices of $H_k(n,p)$, we let $\ell_j(S)$ denote the number of hyperedges that contain exactly j of the vertices of S, and we let $L(S) = \sum_{j=2}^{k}(j-1)\ell_j$.

Lemma 26. *For every $\zeta > 0$, there is $\theta > 0$, such that w.h.p. every S with $|S| \leq \theta n$ has $L(S) < (1+\zeta)|S|$.*

Proof. Rather than working in the $H_k(n,p)$ model with $p = c/n^{k-1}$, where the expected number of edges is $(c/k!)n + o(n)$, it will be convenient to work in the $H^R(n,m)$ model, where exactly $m = (c/k!)n$ edges are selected uniformly, independently and with replacement. Since $m = O(n)$, standard arguments imply that high probability properties in this model transfer to the $H_k(n,p)$ model.

Let $Y_a = Y_a(\zeta)$ denote the number of sets S with $|S| = a$ and $L(S) = (1+\zeta)|S|$. We will bound $\mathbb{E}(Y_a)$ as follows. Define

$$\mathcal{L}_a = \left\{ (\ell_2, ..., \ell_k) : \sum_{j=2}^{k} (j-1)\ell_j \geq (1+\zeta)a \right\}.$$

Choose a vertices and some $(\ell_2, ..., \ell_k) \in \mathcal{L}_a$, pick ℓ_j edges for each j, and then multiply by the probability that each edge chooses (at least) the appropriate number of vertices from S. This yields

$$\begin{aligned}
E(Y_a) &\leq \binom{n}{a} \sum_{(\ell_2,...,\ell_k)\in\mathcal{L}_a} \prod_{j=2}^{k} \binom{m}{\ell_j} \left[\binom{k}{j} \left(\frac{a}{n}\right)^j \right]^{\ell_j} \\
&< \left(\frac{en}{a}\right)^a \sum_{(\ell_2,...,\ell_k)\in\mathcal{L}_a} \left(\frac{a}{n}\right)^{\sum_{j=2}^{k}(j-1)\ell_j} \prod_{j=2}^{k} \frac{(Ja)^{\ell_j}}{\ell_j!} \quad , \quad \text{for some constant } J = J(c,k) > 0 \\
&< \left(\frac{en}{a}\right)^a \left(\frac{a}{n}\right)^{(1+\zeta)a} \prod_{j=2}^{k} \left(\sum_{\ell_j \geq 0} \frac{(Ja)^{\ell_j}}{\ell_j!} \right) \\
&< e^a \left(\frac{a}{n}\right)^{\zeta a} e^{(k-1)Ja} \\
&= \left(\frac{\Delta a}{n}\right)^{\zeta a} \quad , \quad \text{for some constant } \Delta = \Delta(c,k,\zeta) > 0
\end{aligned}$$

Choosing $\theta = \frac{1}{2\Delta}$, it is standard and straightforward to show $\mathbb{E}\left(\sum_{a<\theta n} Y_a\right) = o(1)$. □

The following key lemma will allow us to bound the expected number of large trees T.

Lemma 27. *For every vertex $u \in D_i$, if T is the BFS tree in D_i rooted at u, then*

- *For every i,*
  - *(a) For every $v \in S_i(u)$, $v \neq u$ we have $\deg_{H_i}(v) \leq \deg_{S_i(u)}(v) + r - 2$.*
  - *(b) For every leaf $v \neq u$ of T, $\deg_{H_i}(v) \leq \deg_T(v) + r - 2$.*
- *For any $\delta > 0$, we can choose $i = i(\delta)$ such that*
  - *(c) w.h.p. $\deg_{S_i(u)}(v) = \deg_T(v)$ for all but at most $\delta|S_i(u)| + 2$ vertices $v \in S_i(u)$.*

Proof.

Part (a): Consider $v \neq u$. At the point in the stripping sequence that v will be removed, it will have degree at most $r-1$. One of those $r-1$ edges is in $S_i(u)$, since v has indegree 1. Every other edge of H_i containing v will be removed before v, therefore resulting in a directed edge out of v in $S_i(u)$.

Part (b): Since v is a leaf in T, and $v \neq u$, v was initially deletable. Thus, it has degree at most $r-1$ in H_i and it has degree at least 1 in T.

Part (c): For $2 \leq j \leq k$, let ℓ_j denote the number of hyperedges with j vertices in $S_i(u)$. All vertices of $S_i(u)$, except possibly u itself, are not in the r-core. So, by Lemma 19(a), we know that for any $\theta > 0$ we can select $i = i(\theta)$ such that $|S_i(u)| < \theta n$. If we pick θ sufficiently small in terms of δ, then Lemma 26 implies that w.h.p., $\sum_{j=2}^{k}(j-1)\ell_j < (1+\delta/2)|S_i(u)|$. So

$$\sum_{v\in S_i(u)} \deg_{S_i(u)}(v) = \sum_{j=2}^{k} j\ell_j \leq 2\sum_{j=2}^{k}(j-1)\ell_j < (2+\delta)|S_i(u)|.$$

Now the total T-degree of the vertices in T is $2|S_i(u)| - 2$. So

$$\sum_{v \in S_i(u)} \deg_{S_i(u)}(v) - \deg_T(v) \leq (2+\delta)|S_i(u)| - (2|S_i(u)| - 2) = \delta|S_i(u)| + 2.$$

This proves part (c). □

In the following we will need to take B sufficiently large for various bounds to hold. Let $X_a = X_a(B)$ be the number of BFS trees T in D_B with a vertices. Since we only need to show that there exists some constant $Q > 0$ such that w.h.p. $X_a = 0$ for $a > Q \log n$, in the following we will allow ourselves to assume that a is greater than some sufficiently large constant.

To prove Lemma 24 we first observe that, by Proposition 19, we can assume that H_B is uniformly random conditional on its degree sequence. Since Lemma 24 asserts a property to hold with high probability, it suffices to establish this property in the configuration model for H_B. Moreover, recall that by Proposition 19(b) as B is increased, w.h.p. the degree sequence of H_B tends to that of the core.

Let $v_1, ..., v_a$ be the vertices of T. We first specify $d_i = \deg_T(v_i)$ for each i, noting that these degrees must sum to $2a - 2$. The number of ways to arrange these a vertices into a tree with a specified degree sequence is $(a-2)!/\prod(d_i - 1)!$ and there are a choices for the root, u, of the tree.

Next we choose the vertices of T. Then for each edge of T, we choose a vertex-copy of each of its endpoints. To do so, for each v_i, we choose a copy of v_i for each of the d_i edges in T incident with v_i. If $\deg_{H_B}(v_i) = j$, then there are $j!/(j-d_i)!$ choices for the d_i copies of v_i. Since $\deg_{H_B}(v_i) \geq d_i$, the number of choices corresponding to v_i is at most $\sum_{w:\deg_{H_B}(w) \geq d_i} \deg_{H_B}(w)!/(\deg_{H_B}(w) - d)!$. Applying Lemma 16 with $\theta = \sqrt{\delta}$, and replacing J with ∞, we obtain that this number is at most $(Y(d_i) + \sqrt{\delta})n$ where

$$Y(d) = Y_B(d) = \sum_{j \geq d} \frac{j!}{(j-d)!} \rho_j(B) \ .$$

Note that this bound holds uniformly, even when d_i grows with n.

Furthermore, if $\deg_{S_B(u)}(v_i) = \deg_T(v_i)$, and if $v_i \neq u$, then by Lemma 27(a), $d_i \leq \deg_{H_B}(v) \leq d_i + r - 2$. So in this case we can use $Y'(d_i)$ rather than $Y(d_i)$ where

$$Y'(d) = Y'_B(d) = \sum_{j=d}^{d+r-2} \frac{j!}{(j-d)!} \rho_j(B) \ .$$

Using $Y'(d_i)$ instead of $Y(d_i)$ will be particularly useful when $d_i \leq 2$.

By Lemma 27(c), for any $\delta > 0$ we can take $B = B(\delta) > 0$ sufficiently large, so that we must use $Y(d)$ for at most $\delta a + 2$ vertices $v_i \neq u$, none of which have $\deg_{H_B} = 1$ (since if $\deg_{H_B} = 1$, by Lemma 27(b), we can use Y'). For convenience, we will assume $a > 2/\delta$ so we can take $\delta a + 2 \leq 2\delta a$. Then we overcount by using $Y(d)$ for exactly $2\delta a$ degree 2 vertices, even if the number for which it is required is smaller.

We will upper bound $\mathbb{E}(X_a)$ by using $Y(d)$ for u and for every vertex v_i with $d_i \geq 3$. Let t_1, t_2, t_3 denote the number of vertices v_i for which $d_i = 1, d_i = 2, d_i \geq 3$, respectively. We note that $(Y(\deg_T(u)) + \sqrt{\delta})/(Y'(\deg_T(u)) + \sqrt{\delta}) \leq (Y(1) + \sqrt{\delta})/(Y'(1) + \sqrt{\delta})$. Also, for sufficiently large d, $Y(d)$ is decreasing and so there is a constant d^* such that for all $d \geq 3$, $Y(d) \leq Y(d^*)$. So, if we were to use $Y'(2)$ for every degree 2 vertex, then the overall contribution of the Y, Y' terms would be at most

$$\frac{Y(1) + \sqrt{\delta}}{Y'(1) + \sqrt{\delta}} [(Y'(1) + \sqrt{\delta})n]^{t_1} \cdot [(Y'(2) + \sqrt{\delta})n]^{t_2} \cdot [(Y(d^*) + \sqrt{\delta})n]^{t_3} \ .$$

However, to account for the $2\delta a$ vertices for which we use $Y(2)$, we multiply by the $\binom{t_2}{2\delta a} \leq \binom{a}{2\delta a}$ choices for those vertices, and we multiply by $\Upsilon^{2\delta a}$ where

$$\Upsilon = 2\frac{Y(2)}{Y'(2)} > \frac{Y(2) + \sqrt{\delta}}{Y'(2) + \sqrt{\delta}} \ ,$$

for δ sufficiently small. This brings the overall contribution of the Y, Y' terms to at most:

$$\begin{aligned}
&\frac{Y(1)+\sqrt{\delta}}{Y'(1)+\sqrt{\delta}}[(Y'(1)+\sqrt{\delta})n]^{t_1}\cdot\binom{a}{2\delta a}\Upsilon^{2\delta a}[(Y'(2)+\sqrt{\delta})n]^{t_2}\cdot[(Y(d^*)+\sqrt{\delta})n]^{t_3}\\
= \;& O(n)[(Y'(1)+\sqrt{\delta})n]^{t_1-1}\binom{a}{2\delta a}\Upsilon^{2\delta a}[(Y'(2)+\sqrt{\delta})n]^{t_2}\cdot[(Y(d^*)+\sqrt{\delta})n]^{t_3}.
\end{aligned}$$

Finally, we multiply by the probability that each of the $a-1$ pairs of vertex-copies corresponding to edges of T, lands in a hyperedge of the configuration and divide by the $a!$ rearrangements of the vertices. Because T is a BFS tree rooted at u, each vertex has indegree in T at most 1. Thus, no two edges of T were formed from the same hyperedge. So, we can apply Lemma 12 to the $a-1$ specified pairs of vertex-copies and multiply by $\left(\frac{k-1}{kE}\right)^{a-1}e^{ka^2/(E-a)}$ to get an overall bound, where E is the number of edges in H_B.

Recall that for $c > c^*_{k,r}$, $\mu = \mu(c)$ denotes the larger of the two solutions of $f(\lambda) = c$. By Proposition 19 and Lemma 15 for any $\delta > 0$, we can take B sufficiently large so that

$$\left|kE - \mu\sum_{j\geq r-1}\frac{e^{-\mu}\mu^j}{j!}n\right| \leq \delta n \ .$$

Our key Lemma 20 now yields that by taking B sufficiently large, we can have δ sufficiently small in terms of ζ that various bounds below hold, including

$$\left(\frac{e^{-\mu}\mu^r}{(r-2)!}+\delta\right)\frac{k-1}{kE/n} < 1-\frac{\zeta}{2} \ . \tag{3}$$

By Lemmata 16 and 19, for any $\delta > 0$, we can take B sufficiently large so that $Y'(1) \leq \delta/2$ and $Y'(2) \leq \frac{e^{-\mu}\mu^r}{(r-2)!}+\delta/2$. So, $Y'(1)+\sqrt{\delta}, Y'(2)+\sqrt{\delta}$ are bounded above by δ and $\frac{e^{-\mu}\mu^r}{(r-2)!}+\delta$, respectively. We let $\Psi = 2Y(d^*) > Y(d^*)+\sqrt{\delta}$ for δ sufficiently small.

Putting all this together yields

$$\begin{aligned}
E(X_a) \;\leq\; & O(n)\binom{a}{2\delta a}\Upsilon^{2\delta a}\left(\frac{k-1}{kE}\right)^{a-1}e^{ka^2/(E-a)}\\
&\times\sum_{d_1+\ldots+d_a=2a-2}(\delta n)^{t_1-1}\left[\left(\frac{e^{-\mu}\mu^r}{(r-2)!}+\delta\right)n\right]^{t_2}(\Psi n)^{t_3}\frac{a(a-2)!}{a!\prod_{i=1}^a(d_i-1)!} \qquad (4)\\
\leq\; & O(n/a)e^{ka^2/(E-a)}\left(\frac{1}{(2\delta)^{2\delta}(1-2\delta)^{1-2\delta}}\right)^a\Upsilon^{2\delta a}\left(\frac{k-1}{kE/n}\right)^{a-1}\sum_{d_1+\ldots+d_a=2a-2}\delta^{t_1}\left(\frac{e^{-\mu}\mu^r}{(r-2)!}+\delta\right)^{t_2}\Psi^{t_3}\\
=\; & O(n/a)e^{ka^2/(E-a)}\left(\frac{\Upsilon^{2\delta}}{(2\delta)^{2\delta}(1-2\delta)^{1-2\delta}}\right)^a\left(\frac{k-1}{kE/n}\right)^a\sum_{d_1+\ldots+d_a=2a-2}\delta^{t_1}\left(\frac{e^{-\mu}\mu^r}{(r-2)!}+\delta\right)^{t_2}\Psi^{t_3} \ .
\end{aligned}$$

For δ sufficiently small in terms of ζ,

$$\frac{\Upsilon^{2\delta}}{(2\delta)^{2\delta}(1-2\delta)^{1-2\delta}} < 1+\frac{\zeta}{10} \ .$$

Since this is the degree sequence of a tree, we have $t_1 > t_3$, and so since $\delta < 1$,

$$\begin{aligned}
E(X_a) \;<\; & O(n/a)e^{ka^2/(E-a)}\left(1+\frac{\zeta}{10}\right)^a\\
&\times\sum_{d_1+\ldots+d_a=2a-2}\left(\sqrt{\delta}\frac{k-1}{kE/n}\right)^{t_1}\left(\left[\frac{e^{-\mu}\mu^r}{(r-2)!}+\delta\right]\frac{k-1}{kE/n}\right)^{t_2}\left(\sqrt{\delta}\Psi\frac{k-1}{kE/n}\right)^{t_3} \ .
\end{aligned}$$

We choose δ sufficiently small in terms of ζ so that

$$\sqrt{\delta}\frac{k-1}{kE/n}, \sqrt{\delta}\Psi\frac{k-1}{kE/n} < \frac{\zeta}{100}.$$

This and (3) yield

$$E(X_a) \leq O(n/a)e^{ka^2/(E-a)}\left(1+\frac{\zeta}{10}\right)^a \sum_{d_1+\ldots+d_a=2a-2}\left(1-\frac{\zeta}{2}\right)^{t_2}\left(\frac{\zeta}{100}\right)^{a-t_2}.$$

Now we fix t_2 and count the number of choices for $d_1,...,d_a$. There are $\binom{a}{t_2}$ choices for the values of i with $d_i = 2$. The remaining $a-t_2$ degrees sum to $2a-2-2t_2$. The number of choices for sequences of y non-negative integers that sum to z is $\binom{y+z-1}{y-1}$, so the number of choices for these degrees is bounded by $\binom{2(a-t_2)-3}{a-t_2-1} < 2^{2(a-t_2)-3} < 4^{a-t_2}$. Thus,

$$\begin{aligned}
E(X_a) &\leq O(n/a)e^{ka^2/(E-a)}\left(1+\frac{\zeta}{10}\right)^a \sum_{\ell_2=0}^{a}\binom{a}{t_2}4^{a-t_2}\left(1-\frac{\zeta}{2}\right)^{t_2}\left(\frac{\zeta}{100}\right)^{a-t_2}\\
&= O(n/a)e^{ka^2/(E-a)}\left(1+\frac{\zeta}{10}\right)^a\left(1-\frac{\zeta}{2}+\frac{\zeta}{25}\right)^a\\
&< O(n/a)e^{ka^2/(E-a)}\left(1-\frac{\zeta}{4}\right)^a\\
&< O(n/a)\left(1-\frac{\zeta}{16}\right)^a, \qquad (5)
\end{aligned}$$

where the last inequality holds for all a such that $ka/(E-a) = ka/(\rho n + o(n) - a)$ is sufficiently small that $e^{ka/(E-a)} < 1 + \frac{\zeta}{4}$. Thus, there are constants $Q, \xi > 0$ such that $\mathbb{E}(\sum_{a=Q\log n}^{\xi n} X_a) = o(1)$ and, therefore, w.h.p. there are no trees of size between $Q\log n$ and ξn. Note now that ξ depends only on ζ which, in turn, depends only on c, r. Therefore, we can always select $\delta < \xi$. Recalling that, using Proposition 19(a), we chose B large enough that H_B contains fewer than δn vertices outside of the r-core, this implies that there are no trees of size at least ξn.

□

7.2 Proof of part (b)

Consider any flippable cycle C with vertices $u_1,...,u_\ell$. In our directed graph D, add edges from u_j to u_{j+1} for each j (addition mod ℓ). Thus, $R^+(u_1) = \cup_{j=1}^{\ell} R^+(u_j)$. We modify the arguments from the proof of part (a) for this setting.

Again, we let T be a BFS tree, this time rooted at u_1. Note that since each u_j has degree exactly 2 in the 2-core, it follows that for $\deg_{H_B}(u_j) = \deg_{S_B}(u_j)$ for each j, and for $j = 1, \ell$, $\deg_{S_B}(u_j) = \deg_T(u_j) + 1$. Thus Lemma 27(a) still holds. The only members of C that can be leaves of T are u_1, u_ℓ, so at most one leaf of T violates Lemma 27(b). Finally, by Lemma 23, we can restrict our attention to $\ell = o(n)$. Since $u_1,...,u_\ell$ are the only 2-core vertices in $S_B(U_1)$, we can maintain $|S_B(u_1)| \leq \theta n$ and the proof of Lemma 27(c) still holds.

As in part(a), we bound the expected number of such trees of size a; u_1 is the root and hence plays the role of u from part (a). This time, T has the additional property that there is an edge in D from a vertex of T (i.e. u_ℓ) to u_1. To account for this additional property, We adjust (4) as follows: (i) multiply by the number of choices of one of the $a-1$ other vertices to be u_ℓ; (ii) account for the fact that $\deg_{S_B}(u_i) = \deg_T(u_i)+1$, for $i = 1, \ell$; (iii) choose vertex-copies for the extra edge from u_ℓ to u_1; (iv) adjust the term $\left(\frac{k-1}{kE}\right)^{a-1} e^{ka^2/(E-a)}$ which, by Lemma 12, bounded the probability that the $a-1$ pairs of vertex-copies corresponding to edges of T each landed in a hyperedge of the configuration.

For (ii) and (iii), we use $Y(d(u_j)+1)$ instead of $Y(d(u_j))$ or $Y'(d(u_j))$ for $j = 1, \ell$. Recall that in part (a), we used $Y(\deg(u))$ instead of $Y'(\deg(u))$; since $Y(d(u_1)+1) < Y(\deg(u_1))$, the adjustment for u_1 is a decrease. The adjustment for u_ℓ is an increase of a multiplicative factor of at most $(Y(d(u_\ell)+1)+\sqrt{\delta})/(Y'(\deg(u_\ell))+\sqrt{\delta}) < (Y(d^*)+\sqrt{\delta})/(Y'(1)+\sqrt{\delta}) = O(1)$.

For (iv), the hyperedge containing u_1, u_ℓ is in the 2-core and so is distinct from the other $a-1$ hyperedges. This results in another multiplicative factor of $\frac{k-1}{kE}$ to account for that edge, when applying Lemma 12.

The net result is to multipy $\mathbb{E}(X_a)$ by $O(a/n)$, and so the bound on $\mathbb{E}(X_a)$ in (5) becomes $O(1)\left(1-\frac{\zeta}{16}\right)^a$. Summing over all a yields that the expected number of flippable cycles C such that $|\cup_{u\in C} R^+(u) \cap H_B| > \omega(n)$ is $o(1)$ for any $\omega(n) \to \infty$. Proposition 21 and Lemma 22 complete the proof. □

8 Proof of Lemma 8 below the r-core threshold

As in the case for $c > c^*_{k,r}$, we will carry out a large but fixed number, I, of rounds of the parallel r-stripping process, ending up with a hypergraph H_I. Because we are below the r-core threshold, this will delete all but a very small, albeit linear, number of vertices. Proposition 13 asserts that the remaining hypergraph is uniformly random conditional on its degree sequence. We will determine this degree sequence and apply the technique from [19] to show that the maximum component size in the remaining hypergraph has size $O(\log n)$. Thus, for every v, we must have $|R^+(v) \cap H_I| = O(\log n)$. Proposition 21 and Lemma 22 then implies that $|R^+(v)| = O(\log n)$ as required.

In what follows, we will describe the proof for the case $r = 2$ and $k \geq 3$, as this is all that is required for Theorem 2, and we write $c^*_{k,2} = c^*_k$. The proof for general r is a straightforward adaption.

We first focus on determining the degree sequence, $\rho_0, \rho_1, ...$ of Proposition 13. Let $\mathrm{Po}(\mu)$ denote a Poisson variable with mean μ and recursively define the following quantities:

$$\begin{aligned} \phi(0) &= 1 \\ \lambda_i &= \frac{c\phi_i^{k-1}}{(k-1)!} \\ \phi_{i+1} &= \Pr[\mathrm{Po}(\lambda_i) \geq 1] \\ z_i &= \Pr[\mathrm{Po}(\lambda_{i-1}) \geq 2] \ . \end{aligned}$$

Lemma 28.

$$\rho_1(i) = \Pr[\mathrm{Po}(\lambda_i) = 1] \cdot \Pr\left[\mathrm{Po}\left(\frac{c(z_{i-1}-z_i)}{(k-1)!}\right) \geq 1\right] \tag{6}$$

$$\rho_j(i) = \Pr[\mathrm{Po}(\lambda_i) = j] \ , \ \textit{for } j \geq 2 \ . \tag{7}$$

Proof. For $j \geq 2$, a vertex v has degree j in H_i iff it has exactly j neighbours that survived the first i rounds. As proven, for example, in [18] this occurs with probability that tends to $\Pr[\mathrm{Po}(\lambda_i) = j]$ as $n \to \infty$.

A vertex v has degree 1 in H_i iff it has exactly one neighbour that survived the first i rounds and at least one neighbour that survived the first $i-1$ rounds but not the ith round (otherwise v would have been deleted during round i or earlier). Arguing very similarly to the proof of (7) in [18], it follows that this occurs with probability that tends to the right hand side of (6) as $n \to \infty$. □

A simple adaptation of the proof of the main result of [19] provides a hypergraph version. Proposition 13 and Lemma 16 allow us to apply that hypergraph version to deduce that if

$$(k-1)\sum_{j\geq 1} j(j-1)\rho_j(I) < \sum_{j\geq 1} j\rho_j(I), \qquad \text{i.e.,} \qquad \rho_1(I) > \sum_{j\geq 3}((k-1)j(j-1)-j)\rho_j(I) \ ,$$

then w.h.p. all components of H_I have size $O(\log n)$.

As discussed in [18], if $c < c_k^*$ then $\lim_{i\to\infty} \phi(i) = 0$. (Indeed, it is not hard to see that c_k^* is defined to be the largest c for which the limit of ϕ is 0.) Thus, by taking i sufficiently large, we can take ϕ_i to be arbitrarily small. It will be useful to develop the following asymptotics as $\lambda_{i-1} \equiv \lambda \to 0$:

$$\begin{aligned} z_i &= 1 - e^{-\lambda}(1+\lambda) \\ &= 1 - [1 - \lambda + \lambda^2/2 - O(\lambda^3)](1+\lambda) \\ &= \frac{\lambda^2}{2} + O(\lambda^3) \ . \end{aligned}$$

$$\begin{aligned} \phi_i &= 1 - e^{-\lambda} \\ &= \lambda + O(\lambda^2) \ . \end{aligned}$$

$$\begin{aligned} z_{i-1} &= \frac{\lambda_{i-2}^2}{2} + O(\lambda_{i-2}^3) \\ &= \frac{\phi_{i-1}^2}{2} + O(\phi_{i-1}^3) \\ &= \Theta(\phi_i^{2/(k-1)}) \\ &= \Theta(\lambda^{2/(k-1)}) \ . \end{aligned}$$

From the above, for $k > 2$, it easily follows that for λ sufficiently small, $z_{i-1} - z_i \geq \frac{1}{2} z_{i-1} = \Theta(\lambda^{2/(k-1)})$, and so

$$\rho_1(i) = \lambda_i e^{-\lambda_i} \left(1 - \exp\left(-\frac{c(z_{i-1} - z_i)}{(k-1)!}\right)\right) = \lambda_i e^{-\lambda_i} \times \Theta(\lambda^{2/(k-1)}) \ .$$

At the same time, since $\lambda_i \to 0$, a series expansion easily gives the second equality below

$$\sum_{j\geq 3}((k-1)j(j-1) - j)\rho_j(i) = \sum_{j\geq 3} j(j-2)\Pr[\mathrm{Po}(\lambda_i) = j] = \Theta(\rho_3(i)) = \Theta(\lambda_i^3)e^{-\lambda_i}.$$

Thus, for any $\epsilon > 0$, if i is sufficiently large, we have $\rho_1(i) > (1+\epsilon)\sum_{j\geq 3}((k-1)j(j-1) - j)\rho_j(I)$, since $\lambda_i \lambda^{2/(k-1)} \gg \lambda_i^3$.

9 Proof of Theorem 1

Given a satisfying assignment, we say that a set S of variables is *flippable* if changing the assignment of every variable in S results in another satisfying assignment. A flippable set is *minimal* if it does not contain a flippable proper subset. Note that flippable sets can be characterized in terms of the underlying hypergraph.

Proposition 29. *S is flippable iff every hyperedge contains an even number of members of S.*

Thus, recalling Definition 2, a flippable cycle is a flippable set. Theorem 1 follows from the following.

Lemma 30. *Let H be a random k-uniform hypergraph $H_k(n,p)$, where $p = c/n^{k-1}$. For every $c > c_{k,2}^*$ there exists $\alpha > 0$ such that w.h.p. every minimal flippable set in the 2-core of H either is a flippable cycle or has size at least αn.*

If we could show (deterministically) that the hypergraph induced by any minimal flippable set in a 2-core that is not a flippable cycle is sufficiently dense, then Lemma 30 would follow by a rather standard argument. Unfortunately, there is no useful lower bound on the density, mainly because of the possibility of very long 2-linked paths in S (defined below). Instead, we follow an approach akin to that of [20], forming a graph $\Gamma(S)$ by contracting those long paths, and making use of the fact that $\Gamma(S)$ is dense (Lemma 35). While the

basic idea is similar to [20], the computations are significantly more challenging as we have to carry out the proof in the configuration model.

To prove Lemma 30, we first require a few definitions. Note that these concern any hypergraph, not just a random one.

Definition 31. *Let $\mathcal{H}$ be a k-uniform hypergraph. A* 2-linked path *of a set $S \subseteq V(\mathcal{H})$ is a set of vertices $v_0, ..., v_t \in S$ and hyperedges $e_1, ..., e_t$ such that*

(i) $v_0, ..., v_t$ are all distinct except that possibly $v_0 = v_t$.

(ii) Each e_i contains v_{i-1}, v_i and no other vertices of S.

(iii) $v_1, ..., v_{t-1}$ all have degree 2 in $\mathcal{H}$.

(iv) Each of v_0, v_t either has degree $\neq 2$ in $\mathcal{H}$, or lies in a hyperedge not containing exactly 2 members of S.

We call v_0, v_t the endpoints *of the path and $v_1, ..., v_{t-1}$ its* connecting vertices.

Note that if $v_0 = v_t$ then (iv) implies that $\deg_{\mathcal{H}}(v_0) > 2$ and hence $v_0, ..., v_t$ do not form a flippable cycle. Note also that a loop, i.e., a hyperedge containing a vertex v twice, can yield a 2-linked path with $t = 1$ and $v_0 = v_1 = v$ if $v \in S$, no other vertices of the edge are in S, and $\deg_{\mathcal{H}}(v) > 2$. If $\deg_{\mathcal{H}}(v) = 2$ then v is a flippable cycle.

Definition 32. *We say that $S \subseteq V(\mathcal{H})$ is a* linked set *if (i) S does not contain a flippable cycle as a subset, (ii) no hyperedge of $\mathcal{H}$ contains exactly one element of S and (iii) every hyperedge e of $\mathcal{H}$ with $|e \cap S| = 2$ is in a 2-linked path of S.*

Proposition 33. *Suppose S is a flippable set which does not contain a flippable cycle as a subset. Then S is a linked set.*

Proof. By Proposition 29, we only need to check condition (iii). Consider any hyperedge e with $|e \cap S| = 2$. Either e itself forms a 2-linked path in S, or it is easily seen that e can be extended into such a path, unless e lies in a flippable cycle. □

Remark: It is easy to see that in any Uniquely Extendible CSP, the set of disagreeing variables of any two solutions must be a flippable set. Since Proposition 33 was derived by only considering the underlying hypergraph (and not the specific constraints), it applies to any UE CSP. Therefore, our Theorem 1 extends readily to all UE CSP since its proof amounts to proving that for some constant $\alpha > 0$, all linked sets are either flippable cycles or contain at least αn variables.

Given a linked set, S, we consider the mixed hypergraph $\Gamma(S)$ formed as follows:

(a) The vertices of $\Gamma(S)$ are the endpoints of the 2-linked paths in S along with all vertices of S that do not lie in any 2-linked paths.

(b) There is an edge in $\Gamma(S)$ between the endpoints of each 2-linked path in S.

(c) For every hyperedge e of $\mathcal{H}$ with $|e \cap S| > 2$, $e \cap S$ is a hyperedge of $\Gamma(S)$.

Thus $V(\Gamma(S)) \subseteq S$, and since no hyperedge of C contains exactly one element of S, for every $v \in V(\Gamma(S))$ we have $\deg_{\Gamma(S)}(v) = \deg_{\mathcal{H}}(v)$. Any vertex of S that is not in $\Gamma(S)$ is a connecting vertex of a 2-linked path in S.

Proposition 34. *If S is a non-empty linked set, then $\Gamma(S)$ has at least one vertex.*

Proof. Any vertex of S that is not in $\Gamma(S)$ is a connecting vertex of a 2-linked path in S. The endpoints of that 2-linked path are in $\Gamma(S)$. □

Note that $\Gamma(S)$ contains hyperedges of size between 2 and k. For each $2 \le i \le k$, we define ℓ_i to be the number of i-edges in $\Gamma(S)$.

Lemma 35. *If every vertex in $\mathcal{H}$ has degree at least 2 then $\sum_{i=2}^{k}(i-1)\ell_i \ge (1+\frac{1}{2k})|V(\Gamma(S)|$.*

Proof. As we said above, every $v \in V(\Gamma(S))$ has the same degree in $\Gamma(S)$ as it does in $\mathcal{H}$. Thus $\Gamma(S)$ has minimum degree at least 2. Consider any v of degree 2 in $\Gamma(S)$. Then v has degree 2 in $\mathcal{H}$ and hence cannot be the endpoint of a 2-linked path in S, unless v lies in at least one hyperedge of $\mathcal{H}$ containing more than 2 members of S. It follows that v lies in at least one hyperedge of $\Gamma(S)$ of size greater than 2. Therefore, at most $\sum_{i=3}^{k} i\ell_i < k\sum_{i=3}^{k}\ell_i$ vertices of $\Gamma(S)$ have degree 2, and so letting Z denote the number of vertices with degree at least 3 in $\Gamma(S)$, we have

$$|V(\Gamma(S))| \le Z + k\sum_{i=3}^{k}\ell_i \le k\left(Z + \sum_{i=3}^{k}\ell_i\right).$$

By the handshaking lemma, $\sum_{i=2}^{k} i\ell_i = \sum_v \deg_{\Gamma(S)}(v)$. Therefore,

$$\begin{aligned}
\sum_{i=2}^{k}(i-1)\ell_i &= \frac{1}{2}\sum_v \deg_{\Gamma(S)}(v) + \sum_{i=2}^{k}(i/2-1)\ell_i \\
&\ge \frac{1}{2}\sum_v \deg_{\Gamma(S)}(v) + \frac{1}{2}\sum_{i=3}^{k}\ell_i \\
&= \sum_v 1 + \sum_v \frac{1}{2}(\deg_{\Gamma(S)}(v)-2) + \frac{1}{2}\sum_{i=3}^{k}\ell_i \\
&\ge |V(\Gamma(S))| + \frac{1}{2}Z + \frac{1}{2}\sum_{i=3}^{k}\ell_i \quad , \quad \text{since } \deg_{\Gamma(S)(v)} \ge 2 \text{ for all } v \\
&\ge \left(1+\frac{1}{2k}\right)|V(\Gamma(S))| \ .
\end{aligned}$$

□

Let C be the 2-core of $H = H_k(n,p)$. We will apply Lemma 35 with $\mathcal{H} = C$ to prove:

Lemma 36. *There exists $\alpha > 0$ such that w.h.p. C has no linked set of size less than αn.*

Lemma 30 follows immediately from Lemma 36. The proof of Lemma 36 will be reminiscent of the proof of Lemma 26, but significantly more complicated because (i) we are working in the configuration model and (ii) where we had ℓ_2 2-edges in Lemma 36, we have ℓ_2 2-linked paths here. First, we provide a technical lemma.

Lemma 37. *For any integers a, t, given a set of a vertices in $H = H_k(n,p)$, with $p = c/n^{k-1}$ the probability that their total degree exceeds $tkca$ is at most $(e/t)^{act}$.*

Proof. Given a set A of a vertices, let E_A denote the number of hyperedges containing at least one member of A. The total degree in A is at most kE_A. The number of potential edges in E_A is at most $a\binom{n}{k-1} < an^{k-1}$, and so E_A is dominated from above by $\mathrm{Bin}(an^{k-1}, c/n^{k-1})$ and using $\binom{n}{z} \le (ne/z)^z$ we get

$$\mathbf{Pr}\left[\mathrm{Bin}(an^{k-1}, c/n^{k-1}) > act\right] < \binom{an^{k-1}}{act}\left(\frac{c}{n^{k-1}}\right)^{act} < (e/t)^{act}.$$

□

Proof of Lemma 36. By Corollary 11, we can work in the configuration model. Let $\mathcal{D}$ be the degree sequence of C. Recalling Definition 17, Proposition 18 and our key Lemma 20, we have w.h.p.

(i) $\mathcal{D}$ has total degree $\gamma n + o(n)$, where $\gamma = \mu\Psi_r(\mu)$,

(ii) $\mathcal{D}$ has $\lambda_2 n + o(n)$ vertices of degree 2, where $\lambda_2 = e^{-\mu}\mu^2/2$,

(iii) there exists $\zeta > 0$ such that $2(k-1)\lambda_2 < (1-\zeta)\gamma$.

For each $a \geq 1$, let X_a denote the number of linked sets S in C for which $|\Gamma(S)| = a$ and let $X = \sum_{a=1}^{\alpha n} X_a$. Define

$$\mathcal{L}_a = \left\{ (\ell_2, ..., \ell_k) : \left(1 + \frac{1}{2k}\right) a \leq \sum_{i=2}^{k} (i-1)\ell_i \leq \left(1 + \frac{1}{2k}\right) a + (k-1) \right\}.$$

By Lemma 35, for any linked set S in C of size a, there is some $(\ell_2, ..., \ell_k) \in \mathcal{L}_a$ so that $\Gamma(S)$ contains *at least* ℓ_i i-edges for each i.

To bound $E(X_a)$, we begin by choosing a vertices, $A \subseteq V(C)$ and sum over all $t \geq 0$ of the probability that their total degree in C lies in the range $(tkca, (t+1)kca]$. For each t, we upper bound this last probability by the probability that their total degree in H lies in $(tkca, \infty]$. Moreover, to sum over all subsets $A \subseteq V(C)$ we overcount by summing instead over all such $A \subseteq V(H)$, and using Lemma 37. Of course, if such a set is not a subset of C then the probability of it contributing to X_a is zero. So this provides an upperbound on $\mathbb{E}(X_a)$.

Given A, we sum over all possibilities for the values of $(\ell_2, ..., \ell_k) \in \mathcal{L}_a$. For each $r \geq 2$, we choose ℓ_r r-sets of vertex-copies belonging to vertices of A. If the total degree of A is in $(tkca, (t+1)kca]$ then the number of choices for these ℓ_r $r-sets$ is at most

$$\left(\frac{((t+1)kca)^r}{r!}\right)^{\ell_r} / \ell_r! < \frac{((t+1)kca)^{r\ell_r}}{\ell_r!} \ .$$

Denote the ℓ_2 2-sets as $\{u_1, w_1\}, ..., \{u_{\ell_2}, w_{\ell_2}\}$. For each $i = 1, ..., \ell_2$, we select $j_i \geq 0$, the number of connecting variables in the 2-linked path from u_i to w_i, we choose the j_i degree two connecting variables for that path, and we choose one of the two possible orientations of the vertex-copies of each of those connecting variables. Let $J = j_1 + ... + j_{\ell_2}$, be the number of connecting variables selected. Let $L = \lambda_2 n + o(n)$ be the number of degree 2 vertices in C. Then the total number of choices for the connecting vertices and the orientations of their copies is at most $\prod_{i=1}^{J} 2(L-i+1)$.

Next, we apply Lemma 12 to bound the probability that the $\ell_3 + ... + \ell_k$ sets of size at least 3 all land in hyperedges of the configuration and that for each $i = 1, ..., \ell_2$, the *first* pair in the 2-linked path, i.e., u_i and the first copy of the first of the j_i connecting variables, lands in a hyperedge of the configuration. Note that $\ell_2 + ... + \ell_k \leq \sum_{i=2}^{k}(i-1)\ell_i < 2a + o(n)$. By assuming $a < \alpha n$ for some sufficiently small α, we get $\gamma n - 2a + o(n) > \frac{1}{2}\gamma n$. Therefore, Lemma 12 yields that this probability is at most

$$\begin{aligned} & \exp\left(\frac{k(\ell_2 + ...\ell_k)^2}{\frac{1}{2}\gamma n}\right) \prod_{i=2}^{k} \left(\frac{(k-1)(k-2)...(k-i+1)}{(\gamma n + o(n))^{i-1}}\right)^{\ell_i} \\ = \ & \exp\left(\frac{8ka^2 + o(n)}{\gamma n}\right) \prod_{i=2}^{k} \left(\frac{(k-1)(k-2)...(k-i+1)}{(\gamma n + o(n))^{i-1}}\right)^{\ell_i} \\ < \ & \exp\left(\frac{8ka^2}{\gamma n}\right) \prod_{i=2}^{k} \left(\frac{k}{\gamma n}\right)^{(i-1)\ell_i} \ . \end{aligned}$$

Following the analysis of Lemma 12, we have now exposed $\ell_2 + ... + \ell_k$ hyperedges of the configuration. Let Λ be the number of unmatched vertex-copies remaining. Since $\ell_2 + ... + \ell_k < 2a + o(n)$, we have $\Lambda \geq \gamma n - 2ka + o(n)$. If the other vertex-copies required for the 2-linked paths are still unmatched, then we

continue; else we halt observing that in this case, the set of choices made so far cannot lead to a linked set on the chosen vertices.

There are J pairs of vertex copies that each need to be in a hyperedge of the configuration in order to complete the 2-linked paths. Following the same argument as in Lemma 23, the probability of this happenning is at most $\prod_{i=1}^{J} \frac{k-1}{\Lambda-k(i-1)}$. Applying (iii) above, we obtain the second inequality below

$$\frac{2(k-1)L}{\Lambda} < \frac{2(k-1)\lambda_2 n + o(n)}{\gamma n - 2ka + o(n)} < 1 - \frac{\zeta}{2} \ ,$$

if $a < \alpha n$ for α sufficiently small in terms of γ, λ_2. Since $2(k-1)L \le \Lambda$ and $k \le 2(k-1)$, we have $\frac{2(k-1)(L-(i-1))}{\Lambda-k(i-1)} < 1 - \frac{\zeta}{2}$ for each i, leading to

$$\begin{aligned}
\mathbb{E}(X_a) &< \binom{n}{a} \sum_{t\ge 0} (\frac{e}{t})^{tca} \sum_{\ell_2,\dots,\ell_k \in \mathcal{L}_a} \sum_{j_1,\dots,j_{\ell_2}\ge 0} e^{8ka^2/(\gamma n)} \left(\prod_{i=2}^{k} \frac{((t+1)kca)^{i\ell_i}}{\ell_i!} \right) \\
&\quad \times \left(\prod_{i=2}^{k} \left(\frac{k}{\gamma n} \right)^{(i-1)\ell_i} \right) \left(\prod_{i=1}^{J} \frac{2(k-1)(L-(i-1))}{\Lambda - k(i-1)} \right) \\
&< \left(\frac{en}{a} \right)^a \sum_{t\ge 0} (\frac{e}{t})^{tca} \sum_{\ell_2,\dots,\ell_k \in \mathcal{L}_a} e^{8ka^2/(\gamma n)} \left(\prod_{i=2}^{k} \frac{(kca)^{\ell_i}}{\ell_i!} \left(\frac{k^2 ca}{\gamma n} \right)^{(i-1)\ell_i} (t+1)^{i\ell_i} \right) \\
&\quad \times \sum_{j_1,\dots,j_{\ell_2}\ge 0} (1-\zeta/2)^J \ .
\end{aligned}$$

Since $J = j_1 + \dots + j_{\ell_2}$, we have $\sum_{j_1,\dots,j_{\ell_2}\ge 0} (1-\zeta/2)^J = \left(\sum_{j\ge 0} (1-\zeta/2)^j \right)^{\ell_2} = (2/\zeta)^{\ell_2}$, we get

$$\mathbb{E}(X_a) < \left(\frac{en}{a} \right)^a e^{8ka^2/\gamma n} \sum_{\ell_2,\dots,\ell_k \in \mathcal{L}_a} \left(\frac{k^2 ca}{\gamma n} \right)^{\sum_{i=2}^{k}(i-1)\ell_i} \left(\prod_{i=2}^{k} \frac{(kca)^{\ell_i}}{\ell_i!} \right) (\frac{2}{\zeta})^{\ell_2} \sum_{t\ge 0} (\frac{e}{t})^{tca} (t+1)^{\sum_{i=2}^{k} i\ell_i}.$$

By our choice of $\mathcal{L}_a$

$$\ell_2 \le \sum_{i=2}^{k} (i-1)\ell_i \le (1+\tfrac{1}{2k})a + k - 1,$$

$$\sum_{i=2}^{k} i\ell_i \le 2 \sum_{i=2}^{k} (i-1)\ell_i \le 3a + 2k,$$

we obtain $(2/\zeta)^{\ell_2} < Z_1^a$ and

$$\sum_{t\ge 0} (e/t)^{tca} (t+1)^{\sum_{i=2}^{k} i\ell_i} < \sum_{t\ge 0} (e/t)^{tca} (t+1)^{3a+2k} < \sum_{t\ge 0} \left((e/t)^{tc} (t+1)^{3+2k} \right)^a < Z_2^a,$$

for constants $Z_1 = Z_1(c)$, $Z_2 = Z_2(c)$, since $(e/t)^{tc}(t+1)^{3+2k}$ is decreasing for large t. Also using $a \le n$ we obtain

$$\begin{aligned}
\mathbb{E}(X_a) &< \left(\frac{en}{a} \right)^a e^{8ka/\gamma} (Z_1 Z_2)^a \left(\frac{k^2 ca}{\gamma n} \right)^{\left(1+\frac{1}{2k}\right)a+k-1} \sum_{\ell_2,\dots,\ell_k \ge 0} \prod_{i=2}^{k} \frac{(kca)^{\ell_i}}{\ell_i!} \\
&= O(1) \left(eZ_1 Z_2 e^{8k/\gamma} \left(\frac{k^2 c}{\gamma} \right)^{1+\frac{1}{2k}} \right)^a \left(\frac{a}{n} \right)^{a/2k} \left(\sum_{\ell\ge 0} \frac{(kca)^\ell}{\ell!} \right)^{k-1}
\end{aligned}$$

Applying $\left(\sum_{\ell\geq 0}\frac{(kca)^\ell}{\ell!}\right)^{k-1} = e^{kca(k-1)}$ we get

$$\mathbb{E}(X_a) < O(1)\left(eZ_1Z_2e^{8k/\gamma}(k^2c/\gamma)^{1+\frac{1}{2k}}e^{ck(k-1)}\right)^a\left(\frac{a}{n}\right)^{a/2k} < Y^a\left(\frac{a}{n}\right)^{a/2k},$$

for a constant $Y = Y(\gamma,\lambda_2,\zeta,b,\xi)$ that does not depend on a, so long as $a < \alpha n$ for sufficiently small $\alpha > 0$. This yields $\mathbb{E}(\sum_{a=1}^{\sqrt{n}} X_a) = o(1)$. Moreover, for all α sufficiently small, $\mathbb{E}(X_a) < 2^{-a}$. Therefore, $\mathbb{E}(\sum_{a\geq\sqrt{n}} X_a) = o(1)$ and, thus, $\mathbb{E}(X) = o(1)$. □

10 Acknowledgements

We have become aware that Ibrahimi, Kanoria, Kraning and Montanari concurrently obtained similar results. These papers are independent.

References

[1] D. Achlioptas, P. Beame, and M. Molloy. *A sharp threshold in proof complexity yields lower bounds for satisfiability search.* J. Comput. Syst. Sci., **68** (2), 238–268 (2004).

[2] D. Achlioptas and A. Coja-Oghlan. *Algorithmic barriers from phase transitions.* In Proc. 49th Ann. IEEE Symp. on Foundations of Computer Science (FOCS 08), p 793 - 802.

[3] E.R. Berlekamp, R.J. McEliece, and H.C.A. van Tilborg. *On the inherent intractability of certain coding problems.* IEEE Trans. Inform. Theory **24**, 384386 (1978).

[4] B. Bollobás. *A probabilistic proof of an asymptotic formula for the number of labelled regular graphs.* Europ. J. Combinatorics **1** 311-316 (1980).

[5] A. Coja-Oghlan. *On belief propagation guided decimation for random k-SAT.* In Proc. 22nd SODA (2011) 957-966.

[6] A. Coja-Oghlan and C. Efthymiou. *On independent sets in random graphs.* In Proc. 22nd SODA (2011) 136-144.

[7] V. Chvátal and E. Szemerédi. *Many hard examples for resolution.* J. ACM, **35**(4), 759-768 (1988).

[8] H. Connamacher and M. Molloy. *The exact satisfiability threshold for a potentially intractable random constraint satisfaction problem.* In Proc. 45th of FOCS (2004).

[9] O. Dubois and J. Mandler. *The 3-XORSAT threshold.* In Proc. 43rd FOCS (2002), p 769.

[10] D. Fernholz and V. Ramachandran. *Cores and Connectivity in Sparse Random Graphs.* The University of Texas at Austin, Department of Computer Sciences, technical report TR-04-13 (2004).

[11] M. Guidetti and A.P. Young. *Complexity of several constraint-satisfaction problems using the heuristic classical algorithm WalkSAT.* Phys. Rev. E, **84** (1), 011102, July 2011.

[12] S. Janson and M. Luczak. *A simple solution to the k-core problem.* Random Structures Algorithms **30** (2007) 50 - 62 (2007).

[13] S. Janson, T. Łuczak and A. Ruciński. Random Graphs. Wiley, New York (2000).

[14] J.H.Kim. *Poisson cloning model for random graphs.* arXiv:0805.4133v1

[15] M. Luby, M. Mitzenmacher, M.A. Shokrollahi and D. Spielman. *Analysis of low density codes and improved designs using irregular graphs.* Proceedings of STOC (1998).

[16] M. Mézard, F. Ricci-Tersenghi, and R. Zecchina. *Alternative solutions to diluted p-spin models and XORSAT problems.* J. Stat. Phys. **111**, 505, (2003).

[17] M. Mézard, and A. Montanari. *Information, physics, and computation.* Oxford University Press (2009).

[18] M. Molloy *Cores in random hypergraphs and boolean formulas.* Random Structures and Algorithms **27**, 124 - 135 (2005).

[19] M. Molloy and B. Reed. *A critical point for random graphs with a given degree sequence.* Random Structures and Algorithms **6** 161 - 180 (1995).

[20] M. Molloy and M. Salavatipour. *The resolution complexity of random constraint satisfaction problems.* SIAM J. Comp. **37**, 895 - 922 (2007).

[21] A. Montanari, R. Restrepo and P. Tetali. *Reconstruction and clustering in random constraint satisfaction problems,* submitted (2009).

[22] B. Pittel, J. Spencer and N. Wormald. *Sudden emergence of a giant k-core in a random graph.* J. Comb. Th. B **67**, 111 - 151 (1996).

[23] B. Majewski, N. Wormald, G. Havas and Z. Czech. *A family of perfect hashing methods.* Computer Journal **39** (1996), 547-554.