Electronic Colloquium on Computational Complexity, Report No. 76 (2003)

Testing the independence number of hypergraphs

Michael Langberg

September 7, 2003

Abstract

A k-uniform hypergraph G of size n is said to be ε-far from having an independent set of size ρn if one must remove at least εn^k edges of G in order for the remaining hypergraph to have an independent set of size ρn. In this work, we present a natural *property testing* algorithm that distinguishes between hypergraphs which have an independent set of size $\geq \rho n$ and hypergraphs which are ε-far from having an independent set of size ρn. Our algorithm is natural in the sense that we sample $\simeq c(k)\frac{\rho^{2k}}{\varepsilon^3}$ random vertices of G, and according to the independence number of the hypergraph induced by this sample, we distinguish between the two cases above. Here $c(k)$ depends on k alone (*e.g.* the sample size is independent of n).

1 Introduction

A k-uniform hypergraph is a hypergraph $G = (V, E)$ in which each (hyper) edge is of size exactly k. An independent set I in G, is a subset of vertices that do not include any edges (*i.e.* there does not exist an edge $\{v_1, \ldots, v_k\} \in E$ for which $v_i \in I$ for all $i \in \{1, \ldots, k\}$). The size of the maximum independent set in G is denoted by $\alpha(G)$ (and referred to as the independence number). Consider a k-uniform hypergraph G of (vertex) size n which does not have an independent set of size ρn (*i.e.* $\alpha(G) < \rho n$). Let H be a random subgraph of G of size s (*i.e.* H is the subgraph induced by a random subset of vertices in G of size s). In this work we study the minimal value of s for which $\alpha(H) < \rho s$ with high probability.

In general, if our only assumption on G is that $\alpha(G) < \rho n$, we cannot hope to set s to be smaller than n. Hence, we strengthen our assumption on G, to hypergraphs G which not only satisfy $\alpha(G) < \rho n$ but are also *far* from having an independent set of size ρn (we defer defining the exact notion of "far" until later in this discussion). That is, given a hypergraph G which is *far* from having an independent set of size ρn, we ask for the minimal value of s for which (with high probability) a random subgraph of size s does not have an independent set of size ρs. This question (and many other closely related ones) have been studied in (2-uniform hyper) graphs in [GGR98] under the title of *property testing*.

Property testing Let $\mathcal{C}$ be a class of objects, and $\mathcal{P}$ a property of objects from $\mathcal{C}$. Property testing addresses the problem of distinguishing between elements $c \in \mathcal{C}$ which have the property $\mathcal{P}$ and elements that are *far* from having the property $\mathcal{P}$. The aim is to construct efficient (randomized) distinguishing algorithms that sample the given element c in relatively few places. The notion of property testing was first presented by Rubinfeld and Sudan in [RS96] were the testing of algebraic properties of functions was addressed. Goldreich, Goldwasser, and Ron [GGR98] later initiated the

ISSN 1433-8092

study of combinatorial objects in the context of property testing. In their work they studied (2-uniform) graphs and considered several fundamental combinatorial graph properties related to the independence number, chromatic number, size of maximum cut, and size of the maximum bisection of these graphs. Since, many papers have addressed the notion of property testing, both in the context of functions and in the combinatorial setting (*e.g.* see surveys [Gol98, Ron01, Fis01]).

Property testing of hypergraphs has also been studied in the past. Czumaj and Sohler [CS01] initiated this line of study when analyzing the property of being ℓ colorable. Colorability, and other properties of k-uniform hypergraphs (that can be phrased as a Max-k-CNF formula) were also studied in [AdlVKK02, AS03]. In this work, we consider testing the independence number of hypergraphs. To the best of our knowledge, this property has not been addressed in the past (in the context of hypergraphs).

Testing the independence number Goldreich, Goldwasser, and Ron [GGR98] study property testing of the independence number of (2-uniform) graphs. In [GGR98] a graph G of size n is said to be ε-far from having an independent set of size ρn if any set of size ρn in G has at least εn^2 induced edges. It was shown in [GGR98] that if G is ε-far from having an independent set of size ρn then with high probability a random subgraph of size $s = \frac{c \log(1/\varepsilon)\rho}{\varepsilon^4}$, for a sufficiently large constant c, does not have an independent set of size ρs. The sample size s was later improved in [FLS02] to $c\frac{\rho^4}{\varepsilon^3} \log\left(\frac{\rho}{\varepsilon}\right)$ (again c is a sufficiently large constant). It is not hard to verify that this implies a (two-sided error) property testing algorithm for the independence number of G. Namely, given a graph G, one may sample a random subgraph H of G of size s, and exhaustively compute $\alpha(H)$. On one hand, if G happens to have an independent set of size ρn, then with some constant probability p_1 the independence number of H will be at least ρs (notice that the expected value of $\alpha(H)$ is at least ρs). On the other hand, if G is ε-far from having an independent set of size ρn then, as mentioned above, with probability at most $p_2 < p_1$ it is the case that $\alpha(H) \geq \rho s$.

Our results In this work, we extend the analysis of [FLS02] to hypergraphs. Given a k-uniform hypergraph $G = (V, E)$ of size n, we show that if one must remove at least εn^k edges of G in order for the remaining hypergraph to have an independent set of size ρn (*i.e.* G is far from having an independent set of size ρn) then a random subgraph H of G of size $s \simeq c(k)\frac{\rho^{2k}}{\varepsilon^3}$ satisfies $\alpha(H) < \rho s$ with high probability. Here $c(k)$ depends on k alone, which implies that the sample size s is independent of n. We also show a lower bound for the size of s of value $\simeq c(k)\frac{\rho^{2k-1}}{\varepsilon^2}$ (again following the analysis of [FLS02]).

Definition 1.1. *Let A be a subset of V. Define $E(A)$ to be the number of edges in the hypergraph induced by A.*

Definition 1.2. *Let $\rho < 1$. A k-uniform hypergraph $G = (V, E)$ is said to be $\langle \rho, \varepsilon \rangle$-connected iff every subset A of V of size ρn satisfies $E(A) \geq \varepsilon n^k$ (i.e. the number of edges in the subgraph induced by A is greater than εn^k).*

Theorem 1.3. *Let G be a k-uniform hypergraph. Let H be a random sample of G of size $s \geq c2^k k! \frac{\rho^{2k}}{\varepsilon^3} \log\left(\frac{\rho}{\varepsilon}\right)$ for a large constant c.*

1. *If G has an independent set of size ρn, then with probability $\geq 1/4$ the subgraph H will have an independent set of size ρs.*

2. *If G is $\langle \rho, \varepsilon \rangle$-connected then with probability $\leq 1/20$ the subgraph H will have an independent set of size ρs.*

A few remarks are in place. First of all, in the above theorem we consider only hypergraphs which are k-uniform. Theorem 1.3 can be extended to hypergraphs in which any edge is of size at most k. This follows from the fact that any $\langle \rho, \varepsilon \rangle$-connected hypergraph with edges of size k and smaller contains a subgraph (on the same vertex set) which is k-uniform and $\langle \rho, \varepsilon/2 \rangle$-connected. Secondly, throughout this work we analyze the properties of random subsets H which are assumed to be *small*. Namely, we assume that the value of s (as stated in Theorem 1.3) and the parameters ρ, ε and n satisfy (a) $s < c\sqrt{n}$ and (b) $s < c\rho n$ for a sufficiently small constant c. Finally, the constants $1/4$, and $1/20$ presented above are not tight.

Proof techniques We first give a brief overview of the techniques presented in [FLS02] for testing the independence number of (2-uniform) graphs. Given a graph G with vertex size n which is $\langle \rho, \varepsilon \rangle$-connected, the bulk of the work of [FLS02] addresses the study of the minimal value of s for which a random subgraph H of G satisfies $\alpha(H) < \rho s$ with high probability. This is done by analyzing the probability that a random subset H of G satisfies $\alpha(H) < \rho s$ (as a function of ρ, ε, and the sample size s). The main idea behind their proof is as follows. Given a sample size s, they start by bounding the probability that a random subset R of G of size $\ell \geq \rho s$ is an independent set. As any subset R of H is random in G, one may bound the probability that $\alpha(H) \geq \rho s$ using the standard union bound on all subsets R of H of size $\geq \rho s$. [FLS02], analyze this naive strategy and show that it only suffices to bound the probability of a slightly stronger condition than the condition $\alpha(H) \geq \rho s$. Namely, using this scheme, they bound the probability for which $\alpha(H) \geq 4\rho s$ (instead of exactly ρs). The naive strategy is then enhanced in order to bound the probability that $\alpha(H) > \delta \rho s$ as a function of δ for any $\delta > 1$. This now suffices to bound the probability that $\alpha(H) \geq \rho s$ (using a few additional ideas).

In our work, we follow the line of proof presented in [FLS02]. The application of the techniques presented in [FLS02] for testing the independence number of (2-uniform) graphs to the case of k-uniform hypergraphs involves several difficulties. Roughly speaking, these are overcome by considering for each vertex v in the given graph G, all subsets of V which *share* an edge with v. Specifically, for all vertices v we consider the set system $\mathcal{A}_v$ consisting of sets $\alpha = \{u_1, \ldots, u_i\}$ for which there exists an edge e in G which includes both the vertex v and the set α. Similar ideas have been used in the past in the study of hypergraphs (*e.g.* [AS03]).

Structure In Section 2 we prove the first part of Theorem 1.3. In Section 3 (the main section of this work) we prove the second part of Theorem 1.3. This is done in two steps. In Section 3.1 we analyze the naive strategy discussed above. In Section 3.2 we refine our scheme and obtain Theorem 1.3 (2). Finally, in Section 4 we present the lower bound on the sample size s mentioned previously.

2 Proof of Theorem 1.3 (1)

Theorem 1.3 (1) *Let G be a k-uniform hypergraph of size n with $\alpha(G) \geq \rho n$. Let H be a random sample of G of size s, where s is larger than a sufficiently large constant. With probability $\geq 1/4$ the subgraph H will have an independent set of size ρs.*

Proof (sketch) : Let I be an independent set of G of size ρn. Notice that $H \cap I$ is also an independent set. The expected size of $H \cap I$ is ρs. For sufficiently large s, it is not hard to verify (via the central limit theorem, *e.g.* [Fel66]) that with probability greater than 1/4 the size of the set $H \cap I$ will be greater or equal its expectation. □

3 Proof of Theorem 1.3 (2)

3.1 The Naive scheme

Let $G = (V, E)$ be a $\langle \rho, \varepsilon \rangle$-connected k-uniform hypergraph. In this section we study the probability that a random subset R of V of size ℓ is an independent set. We then use this result to bound the probability that a random subset H of G of size s has a large independent set.

We would like to bound (from above) the probability that R induces an independent set. Let $\{r_1, \ldots, r_\ell\}$ be the vertices of R. Consider choosing the vertices of R one by one, such that at each step the random subset chosen so far is $R_i = \{r_1, \ldots, r_i\}$ and the vertex r_{i+1} is chosen from $V \setminus R_i$. Assume that at some stage R_i is an independent set. We would like to show (with high probability) that after adding the remaining vertices $\{r_{i+1}, \ldots, r_\ell\}$ to R_i, the final set R will not be an independent set.

The vertices in $V \setminus R_i$ that cannot be added to R_i are exactly the vertices v that share an edge with some $k-1$ vertices in R_i. Let $N(R_i) = N_i$ be the set of such vertices in $V \setminus R_i$, and let $I(R_i) = I_i$ be $V \setminus N_i$. Consider the next random vertex $r_{i+1} \in R$. If r_{i+1} is chosen from N_i then it cannot be added to R_i, and we view this round as a success regarding the set R_i. Otherwise, r_{i+1} happens to be in I_i and can be added to R_i. But if the addition of r_{i+1} to R_i happens to add many vertices to N_i, we also view this round as a successful round regarding R_i.

Motivated by the discussion above, we continue with the following definitions. For each subset R_i we define the following weighted set systems of subsets of V. Let $RES(R_i, 1)$ (for *restrict*) be the set of singletons $\{v\}$ that share an edge with vertices in R_i. Namely, $\{v\} \in RES(R_i, 1)$ iff there exists vertices $\{w_1, \ldots, w_{k-1}\}$ in R_i such that there is an edge $\{v, w_1, \ldots, w_{k-1}\}$ in E. $RES(R_i, 1)$ is exactly the set N_i defined above. Define the weight of each element in $RES(R_i, 1)$ as n^{k-2}. Let $RES(R_i, 2)$ be the set of *pairs* of vertices v_1, v_2 which (together) share an edge with vertices in R_i. Namely, $\{v_1, v_2\} \in RES(R_i, 2)$ iff there exists vertices $\{w_1, \ldots, w_{k-2}\}$ in R_i such that there is an edge $\{v_1, v_2, w_1, \ldots, w_{k-2}\}$ in E. Define the weight of each element in $RES(R_i, 2)$ as n^{k-3}. Similarly for each $j \in \{1, \ldots, k-1\}$ let $RES(R_i, j)$ be the set of subsets $\{v_1, \ldots, v_j\}$ of V of size j that share an edge with vertices in R_i. Define the weight of each element in $RES(R_i, j)$ as n^{k-j-1}. Finally let $RES_i = RES(R_i)$ be the union of the sets $RES(R_i, j)$ where $j \in \{1, \ldots, k-1\}$. Let $\|RES(R_i, j)\|$ ($\|RES_i\|$) denote the weight of elements in $RES(R_i, j)$ (RES_i). Notice that $\|RES(R_i, j)\| \leq \binom{n}{j} n^{k-j-1} \leq n^{k-1}$ and that $\|RES_i\| \leq \sum_{j=1}^{k-1} \binom{n}{j} n^{k-j-1} \leq k n^{k-1}$.

Definition 3.1. *Let the normalized degree w.r.t. R_i of a vertex $v \in V$ be the amount on which v restricts upon R_i:*

$$d_v(R_i) = \|RES(R_i \cup \{v\})\| - \|RES(R_i)\| = \|RES(R_i \cup \{v\}) \setminus RES(R_i)\|.$$

In the above notice that $RES(R_i) \subseteq RES(R_i \cup \{v\})$. We call a vertex v in V heavy with respect to R_i (or R_i-heavy for short) if $d_v(R_i) \geq \frac{1}{2^k (k-3)!} \frac{\varepsilon}{\rho} n^{k-1}$.

Each subset R_i of V now defines the following partition (LI_i, HI_i, N_i) of V and the set RES_i. Let RES_i and N_i be as defined as above. Let $I_i = V \setminus N_i$. I_i is now partitioned into two parts: vertices in I_i with *low* normalized degree (w.r.t. R_i), denoted as the set LI_i, and vertices with *high* normalized degree, denoted as HI_i. Namely LI_i is defined to be the ρn vertices of I_i with minimal normalized degree and HI_i is defined to be the remaining vertices of I_i. Ties are broken arbitrarily or in *favor* of vertices in R_i (namely, vertices in R_i are placed in I_i before other vertices of identical degree). If it is the case that $|I_i| \leq \rho n$ then LI_i is defined to be I_i, and HI_i is defined to be empty.

We define the partition corresponding to $R_0 = \phi$ as (LI_0, HI_0, N_0), where LI_0 are the ρn vertices of G of minimal normalized degree, HI_0 are the remaining vertices of G, and N_0 is empty. RES_0 is also defined to be empty.

Notice, using this notation, that the subset R_i is an independent set iff $R_i \subseteq I_i$. Moreover, in this case $R_i \subseteq LI_i$ (all vertices of R_i have normalized degree 0). Furthermore, each vertex r_i in an independent set $R = R_\ell = \{r_1, \ldots, r_\ell\}$ satisfies $r_i \in I_{i-1}$.

We are now ready to bound the probability that a random subset $R = \{r_1, \ldots, r_\ell\}$ of G is independent. Let $R_i = \{r_1, \ldots, r_i\}$, and let (LI_i, HI_i, N_i) be the corresponding partition of V defined by R_i. Consider the case in which R is an independent set. As mentioned above, this happens iff for every i the vertex r_i is chosen to be in $I_{i-1} = LI_{i-1} \cup HI_{i-1}$. We would like to show that this happens with small probability (if ℓ is large enough).

Consider the set RES_i as we proceed in the choice of vertices in R. Initially, the subset RES_0 is of weight 0, and it gets larger and larger as we proceed in the choice of vertices in R. Each vertex in $r_i \in HI_{i-1}$ increases the size of RES_{i-1} substantially, while each vertex in LI_{i-1} may only slightly change the size of RES_{i-1}. In the following, we show that there cannot be many vertices $r_i \in R$ that happen to fall into HI_{i-1}. We thus turn to consider vertices r_i that fall in LI_{i-1} (there are almost ℓ such vertices). The size of LI_i is bounded by ρn. Hence, the probability that $r_i \in LI_i$ is bounded by ρ (by our definitions $R_{i-1} \subseteq LI_{i-1}$ and the vertex r_i is random in $V \setminus R_{i-1}$). This implies that the probability that R is an independent set is roughly bounded by ρ^ℓ. Details follow.

Lemma 3.2. *Let G be a $\langle \rho, \varepsilon \rangle$-connected hypergraph. Let $R = \{r_1, \ldots, r_\ell\}$ be a set in G. The number of vertices r_i which satisfy $r_i \in HI_{i-1}$ is bounded by $t = \frac{2^k k! \rho}{\varepsilon}$.*

Proof : We start with the following claim

Claim 3.3. *Let R_i be as defined above, and let (LI_i, HI_i, N_i) and RES_i be its corresponding partition and set system. Let $I_i = LI_i \cup HI_i$. If G is $\langle \rho, \varepsilon \rangle$-connected then, every vertex in HI_i is R_i heavy.*

Proof : Assume that $LI_i = \rho n$ (otherwise HI_i is empty and the claim holds). By our assumptions on G we have that LI_i induces at least εn^k edges. Let $E(LI_i)$ be the set of these edges, and let m be the size of $E(LI_i)$. To simplify our notation, let $LI = LI_i$, $HI = HI_i$, $N = N_i$ and $R = R_i$. Now that the index i is free, we use it in the following new context. For $i \in \{1, \ldots, k-1\}$ consider the following sequence of weighted i-uniform hypergraphs $H_i = (LI, E_i)$ which all have vertex set LI. The edge set E_i is defined to be all subsets of size i of edges in $E(LI)$ (here and throughout our work we consider a (hyper) edge of size k as a subset of vertices of size k). For example, each edge in $E(LI)$ induces $\binom{k}{2}$ edges in H_2, $\binom{k}{3}$ edges in H_3, and so on. The weight of an edge e_i in H_i is equal to the number of edges e in $E(LI)$ which satisfy $e_i \in e$.

Recall the sets $RES(R,j)$ and $RES(R)$. Our goal is to prove the existence of a vertex $v \in LI$ which is R-heavy (this will imply our assertion). Namely a vertex v which when added to R will significantly increase the weight of $RES(R)$. Formally we are looking for a vertex $v \in LI$ which satisfies $\|RES(R \cup \{v\})\| - \|RES(R)\| \geq \frac{1}{2^k(k-3)!}\frac{\varepsilon}{\rho}n^{k-1}$. Notice that the weight of a subset of vertices with respect to $RES(R)$ may (and usually does) differ from its weight with respect to H_i. To avoid confusion we denote the weight of a subset e with respect to $RES(R)$ as $w_R(e)$. Recall that every edge e of size i in $RES(R)$ has weight $w_R(e) = n^{k-i-1}$. The weight of an edge e in H_i will be marked as $w_{H_i}(e)$.

For each i we now consider the weight (with respect to H_i) of the edges in H_i which are in $RES(R)$. Denote this weight as res_i. Notice that $res_1 = 0$ (by our construction $LI \cap N = \phi$, which implies that singletons $\{v\} \subseteq LI$ are not in $RES(R)$). We consider the following cases.

Case 1: We start by assuming that $res_{k-1} \leq (k-1)m$. As the total weight of edges in H_{k-1} is km, we conclude that the weight of edges in H_{k-1} that are not in $RES(R)$ is at least m. Let v be a vertex in LI. We say that an edge e in H_{k-1} is a v-edge if there exists an edge in $E(LI)$ consisting of the union of e and $\{v\}$. Notice that an edge e in H_{k-1} of weight $w_{H_{k-1}}(e)$ is actually a v edge for $w_{H_{k-1}}(e)$ distinct vertices v. As each edge in H_{k-1} is a v-edge for some vertex v in LI, we have the existence of a vertex v in LI with at least $\frac{m}{\rho n}$ (distinct) v-edges that are not in $RES(R)$. The weight of each such v-edge in $RES(R)$ is 1. Furthermore, each such v-edge appears in $RES(R \cup \{v\})$ implying that

$$\|RES(R \cup \{v\})\| - \|RES(R)\| \geq \frac{m}{\rho n} \geq \frac{\varepsilon}{\rho}n^{k-1}$$

which in turn implies the v is R-heavy.

Cases 2 to $k-1$: Starting at $i = k-1$ and iteratively continuing until $i = 2$ consider the following cases. From the previous step, we may assume that $res_i > \frac{2m}{2^{k-i}}\frac{(k-1)}{(k-i)!}$. We now also assume that $res_{i-1} \leq \frac{m}{2^{k-i}}\frac{(k-1)}{(k-i+1)!}$. Consider an edge $e = (v_1, \ldots, v_i)$ in H_i of weight $w_{H_i}(e)$ which contributes to res_i. Each such edge induces i edges in H_{i-1}: $\{e'_1, \ldots, e'_i\}$ (each edge obtained by removing one vertex from e). We are interested in bounding (by below) the weight of edges e as above with corresponding edges e'_j which are not in $RES(R)$ (here $j \in \{1, \ldots, i\}$).

By our construction, the weight of any edge e' in H_{i-1} equals the number of edges in G that include e'. It is not hard to verify that this equals $\frac{1}{k-i+1}$ times the weight of edges in H_i which include e'. We conclude that the weight of edges e in H_i which are in $RES(R)$ with some corresponding edge e' in H_{i-1} which is also in $RES(R)$ is bounded by $res_{i-1}(k-i+1)$. This leaves us with

$$res_i - (k-i+1)res_{i-1} > \frac{m}{2^{k-i}}\frac{(k-1)}{(k-i)!}$$

edges e which contribute to res_i for which e'_j for $j \in \{1, \ldots, i\}$ are not in $RES(R)$.

We now conclude the existence of a vertex $v \in LI$ which is adjacent to (*i.e.* is included in) at least the weight of $\frac{im}{2^{k-i}\rho n}\frac{(k-1)}{(k-i)!}$ edges in H_i which appear in $RES(R)$ such that their corresponding edges in H_{i-1} are not in $RES(R)$. The weight of each edge in H_i is bounded by $\binom{\rho n}{k-i} \leq (\rho n)^{k-i}$. Thus there are at least $\frac{im}{2^{k-i}(\rho n)^{k-i+1}}\frac{(k-1)}{(k-i)!}$ distinct edges in H_i adjacent to v with corresponding edges in H_{i-1} that are not in $RES(R)$.

We will now show that v is R-heavy. Consider one of the distinct edges $e = (v, v_1, \dots, v_{i-1})$ as discussed above. By our assumption the set $\{v_1, \dots, v_{i-1}\}$ is not in $RES(R)$. The edge e appears in $RES(R)$ implying that there exists vertices $\{w_1, \dots, w_{k-i}\}$ (in R) such that

$$(v, v_1, \dots, v_{i-1}, w_1, \dots, w_{k-i})$$

is an edge in the original hypergraph G. This in turn implies that $\{v_1, \dots, v_{i-1}\}$ will be included in $RES(R \cup \{v\})$. As each set of size $i-1$ in $RES(R)$ has weight $w_R = n^{k-i}$, and there are at least $\frac{im}{2^{k-i}(\rho n)^{k-i+1}} \frac{(k-1)}{(k-i)!}$ distinct edges of interest adjacent to v, we conclude that

$$\|RES(R \cup \{v\})\| - \|RES(R)\| \geq \frac{1}{2^{k-i}} \frac{(k-1)}{(k-i)!} \frac{\varepsilon}{\rho} n^{k-1}.$$

□

Now to prove our lemma, consider the subsets $R_i = \{r_1, \dots, r_i\}$ and their corresponding partitions (LI_i, HI_i, N_i). Let $I_i = LI_i \cup HI_i$. We would like to bound the number of vertices r_i that are in HI_{i-1}. Consider a vertex r_i in HI_i. By Claim 3.3, its normalized degree w.r.t. R_i is at least $\frac{1}{2^k(k-3)!} \frac{\varepsilon}{\rho} n^{k-1}$. $RES(\phi)$ is initially empty, and for any i $RES(R_i)$ is of weight at most $\sum_{i=2}^{k-1} \binom{n}{i} n^{k-i-1} \leq kn^{k-1}$. Each vertex $r_i \in HI_{i-1}$ increases $\|RES(R_{i-1})\|$ by at least $\frac{1}{2^k(k-3)!} \frac{\varepsilon}{\rho} n^{k-1}$. We conclude that there are at most $\frac{2^k k! \rho}{\varepsilon}$ vertices r_i in R which are in HI_{i-1}. □

Theorem 3.4. *Let G be a $\langle \rho, \varepsilon \rangle$-connected hypergraph. Let t be as in Lemma 3.2. Let $\ell \geq 2t$. The probability that ℓ random vertices of G induce an independent set is at most*

$$\rho^\ell \left(\frac{e\ell}{t\rho} \right)^t$$

Proof : Let $R = \{r_1, \dots, r_\ell\}$ be a set of ℓ random vertices. As mentioned previously, the probability that $r_i \in LI_{i-1}$ is at most ρ. This follows from the fact that (1) The size of LI_{i-1} is at most ρn, (2) $R_{i-1} \subseteq LI_{i-1}$ (by our definitions), and (3) The vertex r_i is random in $V \setminus R_{i-1}$.

Now in order for R to be an independent set, every vertex r_i of R must be in the set I_{i-1}. Furthermore, by Lemma 3.2 all but t vertices r_i of R must satisfy $r_i \in LI_{i-1}$. Hence, the probability that R is an independent set is at most

$$\binom{\ell}{t} \rho^{\ell-t} \leq \left(\frac{\ell e}{t} \right)^t \rho^{\ell-t} = \rho^\ell \left(\frac{e\ell}{t\rho} \right)^t.$$

□

Let δ be a large constant. We now use Theorem 3.4 to bound the probability that a random subset H of G of size s has an independent set of size $> \delta\rho s$. The result is the following Corollary 3.5. In Section 3.2 we refine our proof techniques and *get rid* of the parameter δ. That is, we bound the probability that a random subset H of G of size s has an independent set of size $\geq \rho s$.

Corollary 3.5. *Let G be a $\langle \rho, \varepsilon \rangle$-connected hypergraph. Let t be as in Lemma 3.2. Let H be a random sample of G of size s. Let $\delta \geq e^2$, and let c be a sufficiently large constant. If $s \geq ct\frac{\log(1/\rho)}{\rho}$ then the probability that $\alpha(H) > \delta\rho s$ is at most $\left(\frac{\varepsilon}{\delta}\right)^{\Omega(\delta\rho s)}$.*

Proof : Let $\ell = \delta\rho s$. Using Theorem 3.4 and the fact that a subset R of H is random in G, the probability that there is an independent set R in H of size k is at most

$$\binom{s}{\ell}\rho^{\ell}\left(\frac{e\ell}{t\rho}\right)^{t} \leq \left[\left(\frac{e}{\delta}\right)^{\frac{\ell}{t}}\frac{e\ell}{t\rho}\right]^{t} \leq \left(\frac{e}{\delta}\right)^{\Omega(\ell)}.$$

In the last inequality we use the fact that $\frac{\ell}{t}$ is greater than $c\log(1/\rho)$, and $\frac{e}{\delta}$ is smaller than $1/e$. □

3.2 An enhanced analysis

Let $G = (V, E)$ be a $\langle\rho, \varepsilon\rangle$-connected hypergraph, and let $H = \{h_1, \ldots, h_s\}$ be a set of random vertices of size s in V. In the previous section we presented a bound on the probability that $\alpha(H) > \delta\rho s$ for large constant values of δ. In this section we enhance our analysis, and bound the probability that $\alpha(H) \geq \rho s$ (namely, we get rid of the additional parameter δ).

Recall our proof technique from Section 3.1. We started by analyzing the probability that a subset R of H of size ℓ is an independent set. Afterwards we bounded the probability that $\alpha(H) > \delta\rho n$ by using the standard union bound on all subsets R of H of size greater than $\ell = \delta\rho n$. In this section we enhance the first part of this scheme by analyzing the probability that a subset R of H of size ℓ is a *maximum* independent set in H (rather than just an independent set of H). Then, as before, using the standard union bound on all large subsets R of H, we bound the probability that $\alpha(H) \geq \rho s$. We show that taking the maximality property of R into account will suffice to prove Theorem 1.3 (2).

Let $H = \{h_1, \ldots, h_s\}$ be s random vertices in G. We would like to analyze the probability that a given subset R of H of size ℓ is a *maximum* independent set. Recall (Section 3.1), that the probability that R is an independent set is bounded by approximately ρ^{ℓ}. An independent set R is a maximum independent set in H only if adding any other vertex in H to R will yield a set which is no longer independent. Let $R = R_\ell$ be an independent set, and let $(LI_\ell, HI_\ell, N_\ell)$ be the partition (as defined in Section 3.1) corresponding to R. Consider an additional random vertex h from H. The probability that $R \cup h$ is no longer an independent set is approximately $|N_\ell|/n$ (here we assume that $|R|$ is small compared to n). The probability that for every $h \in H \setminus R$ the subset $R \cup \{h\}$ is no longer independent is thus $\simeq (|N_\ell|/n)^{s-\ell}$. Hence, the probability that a given subset R of H of size ℓ is a *maximum* independent set is bounded by approximately $\rho^{\ell}\,(|N_\ell|/n)^{s-\ell}$. This value is substantially smaller than ρ^{ℓ} iff $|N_\ell|$ is substantially smaller than n. We conclude that it is in our favor to somehow ensure that $|N_\ell|$ is not too large. We do this in an artificial manner.

Let $R = \{r_1, \ldots, r_\ell\}$ be an independent set, let $R_i = \{r_1, \ldots, r_i\}$, let (LI_i, HI_i, N_i) be the partition (as defined in Section 3.1) corresponding to R_i and let RES_i be the set system corresponding to R_i. Roughly speaking, in Section 3.1, every time a vertex r_i was chosen, the set system RES_i and N_i was updated. If r_i was chosen in HI_{i-1}, then RES_{i-1} increased substantially and N_{i-1} potentially also grew substantially, and if r_i was chosen in LI_{i-1}, both RES_{i-1} and N_{i-1} only slightly changed. We would like to change the definition of the partition (LI_i, HI_i, N_i) and of RES_i corresponding to R_i as to ensure that N_i is always substantially smaller than n. This cannot be done unless we relax the definition of N_i. Recall that N_i was defined (in Section 3.1) to be the set of vertices v which share an edge with vertices in R_i. Specifically, N_i in the set of vertices v for which there exists vertices $\{w_1, \ldots, w_{k-1}\}$ in R_i such that there is an edge $\{v, w_1, \ldots, w_{k-1}\}$ in E. In this section N_i will only include a subset of these vertices (a subset which is substantially smaller than n). Namely, in our new definitions RES_{i-1} and N_{i-1} will be changed only if r_i was chosen in HI_{i-1}. In the case in which $r_i \in LI_{i-1} \cup N_{i-1}$, we do not change N_{i-1} at all. As we will see, such a definition will imply that $|N_i| \leq (1-\rho)s$, which will now suffice for our proof.

A new partition and set system Let $H = \{h_1, \ldots, h_s\}$ be a subset of V. Let $R_i = \{r_1, \ldots, r_i\}$ be a subset of H of size i. In the previous section, the subset R_i defined a partition (LI_i, HI_i, N_i) of V and a set system RES_i. In this section, for each i we will define a subset $\hat{R}_i$ of R_i, and a new partition and set system. The new partition and set system corresponding to R_i will be defined similarly to those defined in the previous section with the exception that $\hat{R}_i$ will play the role that R_i played previously. The set $\hat{R}_i$, the new partition (LI_i, HI_i, N_i), and the set system RES_i are defined as follows (as before, let $I_i = LI_i \cup HI_i$).

1. Initially $R_0 = \hat{R}_0 = \phi$, LI_0 is the ρn vertices in V of minimal normalized degree w.r.t. $\hat{R}_i$, $HI_0 = V \setminus LI_0$, and $N_0 = \phi$. In the above, ties are broken by an assumed ordering on the vertices in V. $RES_0 = RES(\hat{R}_0)$ is defined to be empty.
2. Let $\hat{R}_i$, (LI_i, HI_i, N_i) and RES_i be the sets corresponding to R_i, let r_{i+1} be a new random vertex. Let $R_{i+1} = R_i \cup \{r_{i+1}\}$, we now define the sets $\hat{R}_{i+1}$, $(LI_{i+1}, HI_{i+1}, N_{i+1})$ and RES_{i+1}. Let $N(r_{i+1})$ be the set of vertices which share an edge with vertices in $\hat{R}_i \cup r_{i+1}$. We consider the following cases:
   - If $r_{i+1} \in LI_i$ or $r_{i+1} \in N_i$ then the sets corresponding to R_{i+1} will be exactly those corresponding to R_i. Namely, $\hat{R}_{i+1} = \hat{R}_i$, $LI_{i+1} = LI_i$, $HI_{i+1} = HI_i$, $N_{i+1} = N_i$, and RES_{i+1} will be defined as RES_i.
   - If $r_{i+1} \in HI_i$ then we consider two sub-cases:
     - If $|N_i \cup N(r_{i+1})| \leq (1-\rho)n$, then $\hat{R}_{i+1} = \hat{R}_i \cup \{r_{i+1}\}$, $RES_{i+1} = RES(\hat{R}_{i+1})$, and LI_{i+1}, HI_{i+1}, N_{i+1}, are defined as in Section 3.1. Namely, $N_{i+1} = N_i \cup N(r_{i+1})$. I_{i+1} is defined to be $V \setminus N_{i+1}$. LI_{i+1} is defined to be the ρn vertices of I_{i+1} with minimal normalized degree w.r.t. $\hat{R}_{i+1}$. Finally, HI_{i+1} is defined to be the remaining vertices of I_{i+1}. Ties are broken by the assumed ordering on V.
     - If $|N_i \cup N(r_{i+1})| > (1-\rho)n$, then let $\hat{N}(r_{i+1})$ be the first (according to the assumed ordering on V) $(1-\rho)n - |N_i|$ vertices in $N(r_{i+1})$ and set $N_{i+1} = N_i \cup \hat{N}(r_{i+1})$. Furthermore, set LI_{i+1} to be the remaining ρn vertices of G, and HI_i to be empty. Notice that in this case $|N_{i+1}|$ is of size exactly $(1-\rho)n$. Finally, let $\hat{R}_{i+1} = \hat{R}_i$ and $RES_{i+1} = RES_i$.

A few remarks are in place. First of all it is not hard to verify that the definition above implies

Claim 3.6. *Let $i \in \{1, \ldots, \ell\}$. The sets corresponding to R_i as defined above satisfy (a) $I_i \subseteq I_{i-1}$. (b) $N_{i-1} \subseteq N_i$. (c) $|N_i| \leq (1-\rho)n$. (d) $|LI_i| = \rho n$. (e) The set LI_i is the ρn vertices of minimal normalized degree in I_i (w.r.t. $\hat{R}_i$).*

Secondly, due to the iterative definition of our new partition, the sets corresponding to the subsets R_i depend strongly on the specific ordering of the vertices in R_i. Namely, in contrast to the partitions (set systems) used in Section 3.1, a single subset R with two different orderings may yield two different partitions (set systems). For this reason, in the remainder of this section, we will assume that the vertices of H are chosen one by one. This will imply an ordering on H and on any subset R of H. The partitions we will study will correspond to these orderings only.

Finally, in Section 3.1, an (ordered) subset $R = \{r_1, \ldots, r_\ell\}$ was independent iff $\forall i \quad r_i \in I_{i-1}$ (according to the definition of I_{i-1} appearing in Section 3.1). In this section, if R is independent then it still holds that $\forall i \quad r_i \in I_{i-1}$. However, it may be the case that $\forall i \quad r_i \in I_{i-1}$ but R is not an independent set. In the remainder of this section, we call ordered subsets R for which $\forall i \quad r_i \in I_{i-1}$ *free* sets. We analyze the probability that a random ordered subset H of V of size s does not have any free sets of size larger then ρs. This implies, that H does not include any independent sets of size ρs.

Definition 3.7. *An ordered subset $R_i = \{r_1, \ldots, r_i\}$ is said to be free if it is the case that $r_j \in I_{j-1}$ for all $j \leq i$.*

Proposition 3.8. *Let $H = \{h_1, \ldots, h_s\}$ be an ordered set of vertices in a $\langle \rho, \varepsilon \rangle$-connected hypergraph G. If $\alpha(H) > \rho s$ then the maximum free set in H (w.r.t. the ordering implied by H) is of size $> \rho s$.*

Proof : Let I be an independent set of size $> \rho s$ in H. It is not hard to verify that I (under the ordering implied by H) is a free set. □

Proposition 3.8 implies that to prove Theorem 1.3 (2) it suffices to analyze the maximum free set $R \subseteq H$. Moreover, the only ordered subsets R that we need to consider are those ordered by the ordering implied by H. We now turn to prove Theorem 1.3 (2). Roughly speaking, we start by analyzing the probability that a random subset R is a free set. We then analyze the probability that a given subset R in H is a maximum free set. Finally, we use the union bound on all subsets R of H of size $\geq \rho s$ to obtain our results.

In the remainder of this section, we will assume that the subset H is chosen from G randomly *with repetitions*. That is H is a random multi-set of size s. Our results (with minor modifications) apply also to the case in which H is a random subset of G (and not a multi-set) if the size of H is not very large (here we assume that $|H| << \sqrt{n}$). As in such cases, a set H of size s which is randomly chosen from V with repetitions will not include the same vertex twice (with high probability).

We start by stating the following lemmas which are analogous to Lemma 3.2 and Theorem 3.4 from Section 3.1. The main difference between the lemmas below (and their proof), and those of the previous section is in the definition of the partition (LI_i, HI_i, N_i),the set system RES_i, and in the fact that they address free sets instead of independent sets. Proof of the lemmas is omitted.

Lemma 3.9. *Let G be a $\langle \rho, \varepsilon \rangle$-connected hypergraph. Let $R = \{r_1, \ldots, r_\ell\}$ be an ordered set in G of size ℓ. The number of vertices r_i which satisfy $r_i \in HI_{i-1}$ is bounded by $t = \frac{2^k k! \rho}{\varepsilon}$.*

Lemma 3.10. *Let G be a $\langle \rho, \varepsilon \rangle$-connected hypergraph. Let t be as in Lemma 3.9. Let $\ell \geq 2t$. Let $R = \{r_1, \ldots, r_\ell\}$ be ℓ random vertices of G. The probability that R induces a free set is at most*

$$\rho^\ell \left(\frac{e\ell}{t\rho} \right)^t$$

We now address the probability that a random subset R of H is a maximum free set. We will then use the union bound on all subsets R of H of size $\geq \rho s$ to obtain our results.

Lemma 3.11. *Let G be a $\langle\rho,\varepsilon\rangle$-connected hypergraph. Let t be as in Lemma 3.9. Let $\ell \geq 2t$. Let H be an ordered random sample of G of size $s \geq \ell$. The probability that a given subset R of H of size ℓ is a maximum free set is at most*

$$\rho^{\ell}\left(\frac{e\ell}{t\rho}\right)^{t}(1-\rho)^{s-\ell}$$

Proof : Let $R = \{r_1, \ldots, r_\ell\}$ (ordered by the ordering induced by H). The set R is a maximum free set in H only if (a) R is free, and (b) For each vertex $h \in H$ which is not in R, the ordered set $R^+ = \{r_1, \ldots, r_j, h, r_{j+1}, \ldots, r_\ell\}$ is not free. Here the index j is such that r_j appears before h in the ordering of H, and r_{j+1} appears after h (*i.e.* R^+ is ordered according to the ordering of H).

The probability that R is free has been analyzed in Lemma 3.10. It is left to analyze the probability that R^+ is not free for every vertex $h \notin R$, given that R is free. Consider a vertex $h \in H$ which is not in R, and let $R^+ = \{r_1, \ldots, r_j, h, r_{j+1}, \ldots, r_\ell\}$.

Claim 3.12. *Let $R = \{r_1, \ldots, r_\ell\}$ be a free set. Let $R^+ = \{r_1, \ldots, r_j, h, r_{j+1}, \ldots, r_\ell\}$. Let the partition corresponding to $R_j = \{r_1, \ldots, r_j\}$ be (LI_j, HI_j, N_j), and the set system corresponding to R_j be RES_j If $h \in LI_j$ then R^+ is also a free set.*

Proof : We will use the following notation. Let $R_i = \{r_1, \ldots, r_i\}$ denote the first i vertices of R, let $\hat{R}_i$, (LI_i, HI_i, N_i), and RES_i be its corresponding sets. For $i > j$, let $R_i^+ = \{r_1, \ldots, r_j, h, r_{j+1}, \ldots, r_i\}$ denote the first $i + 1$ vertices of R^+, and let $\hat{R}_i^+$, (LI_i^+, HI_i^+, N_i^+), and RES_i^+ be its corresponding sets. Finally let R_h^+ denote the subset $\{r_1, \ldots, r_j, h\}$ and $\hat{R}_h^+$, (LI_h^+, HI_h^+, N_h^+), and RES_h^+ be its corresponding sets.

We would like to prove that R^+ is free. That is, we would like to show (a) that $r_i \in I_{i-1}$ for each $i \leq j$, (b) that $h \in I_j$, (c) that $r_{j+1} \in I_h^+$, and (d) that $r_i \in I_{i-1}^+$ for $i \geq j + 2$. Recall that R is free and thus $r_i \in I_{i-1}$ for all $i \in \{1, \ldots, \ell\}$.

The first assertion follows from the fact that the first j vertices of R and R^+ are identical. The second follows from the assumption that $h \in LI_j$. For the third assumption, notice (as $h \in LI_j$) that the sets corresponding to $R_h^+ = \{r_1, \ldots, r_j, h\}$ are equal to the sets corresponding to $R_j = \{r_1, \ldots, r_j\}$. This follows from our definitions. As $r_{j+1} \in I_j$ we conclude that $r_{j+1} \in I_h^+$.

For the final assertion, observe that for any $i \geq j+1$, the sets corresponding to R_i^+ are equal to the sets corresponding to R_i. This can be seen by induction (on i). We start with the sets corresponding to R_{j+1} and R_{j+1}^+. The sets corresponding to R_{j+1} are defined uniquely by the sets corresponding to R_j and the vertex r_{j+1}. Similarly, the sets corresponding to R_{j+1}^+ are defined uniquely by the sets corresponding to R_h^+ and the vertex r_{j+1}. As the sets corresponding to R_h^+ and R_j are equal we conclude that the same hold for the sets corresponding to R_{j+1} and R_{j+1}^+. The inductive step is done similarly. □

Claim 3.12 implies that the probability that $R^+ = \{r_1, \ldots, r_j, h, r_{j+1}, \ldots, r_\ell\}$ is not free given that R is free is at most $(1-\rho)$ (recall that the set LI_j is of size exactly ρn). This holds independently for every vertex h in $H \setminus R$. We conclude that the probability that R is a maximum free subset of H is at most the probability that R is free times $(1-\rho)^{s-\ell}$. □

We now turn to analyze the probability that a random ordered subset H of G of size s has a free set of size larger than ρs. We follow the line of analysis given in Section 3.1 and analyze the probability that H has a free set of size larger than $\delta\rho s$ for any $\delta > 1$. We then get rid of the factor δ to obtain our main theorem of this section.

Corollary 3.13. *Let G be a $\langle\rho,\varepsilon\rangle$-connected hypergraph. Let t be as in Lemma 3.9. Let H be a random sample of G of size s. Let $\delta > 1$, and let c be a sufficiently large constant. Let $\Gamma = \ln\delta - \frac{\delta-1}{\delta}$. If $s \geq \frac{ct}{\rho\Gamma}(\log(1/\rho) + \log(1+1/\Gamma))$ then the probability that H has a free set of size $> \delta\rho s$ is at most*

$$s\left(\frac{1}{e^{\Gamma}}\right)^{\Omega(\delta\rho s)}$$

Proof : Let $\ell = \delta\rho s$, let $\delta' > \delta$ and let $\ell' = \delta'\rho s > \ell$. Using Lemma 3.11, the probability that there is a maximum free set R in H of size $> l$ is at most

$$\begin{aligned}
\sum_{\ell'>\ell}\binom{s}{\ell'}\rho^{\ell'}\left(\frac{e\ell'}{t\rho}\right)^t(1-\rho)^{s-\ell'} &\leq \sum_{\ell'>\ell}\left(\frac{e\ell'}{t\rho}\right)^t\frac{s^s(1-\rho)^{s-\ell'}}{\ell'^{\ell'}(s-\ell')^{s-\ell'}}\rho^{\ell'} \\
&\leq \sum_{\ell'>\ell}\left(\frac{e\ell'}{t\rho}\right)^t\frac{e^{\ell'\frac{\delta'-1}{\delta'}}}{\delta'^{\ell'}} \\
&= \sum_{\ell'>\ell}\left[\frac{e\ell'}{t\rho}\left(\frac{1}{e^{\ln\delta'-\frac{\delta'-1}{\delta'}}}\right)^{\frac{\ell'}{t}}\right]^t \\
&\leq \sum_{\ell'>\ell}\left[\frac{e\ell'}{t\rho}\left(\frac{1}{e^{\Gamma}}\right)^{\frac{\ell'}{t}}\right]^t \\
&\leq \sum_{\ell'>\ell}\left(\frac{1}{e^{\Gamma}}\right)^{\Omega(\ell')} \leq s\left(\frac{1}{e^{\Gamma}}\right)^{\Omega(\ell)}
\end{aligned}$$

We use the facts that $\frac{\ell'}{t}$ is greater than both $c\frac{1}{\Gamma}\log(1+1/\Gamma)$ and $c\frac{1}{\Gamma}\log(1/\rho)$ for a sufficiently large constant c. □

It is left to get rid of the additional parameter δ of Corollary 3.13 (namely to analyze the probability that $\alpha(H) \geq \rho s$).

Lemma 3.14. *If G is $\langle\rho,\varepsilon\rangle$-connected then G is also $\left\langle\rho(1-\frac{\varepsilon}{\rho^k}), \frac{\varepsilon}{2}\right\rangle$-connected.*

Proof : Let A be some subset of G of size $\rho(1-\frac{\varepsilon}{\rho^k})n$. Let A^c be any set in $V\setminus A$ of size $\frac{\varepsilon}{\rho^k}\rho n$. It is known that the number of edges induced by the set $A\cup A^c$ is at least εn^k (notice that $|A\cup A^c| = \rho n$). The number of edges (in $A\cup A^c$) adjacent to vertices in A^c is bounded by $\binom{\rho n}{k-1}\frac{\varepsilon}{\rho^k}\rho n \leq \frac{\varepsilon}{2}n^k$. Hence, the number of edges induced by vertices in A is at least $\frac{\varepsilon}{2}n^k$ implying our assertion. □

Theorem 1.3 (2) *Let G be a $\langle\rho,\varepsilon\rangle$-connected hypergraph. Let t be as in Lemma 3.9. Let H be a random sample of G of size s. Let c be a sufficiently large constant. If $s \geq ct\frac{\rho^{2k-1}}{\varepsilon^2}\log\left(\frac{\rho}{\varepsilon}\right)$ then the probability that H has an independent set of size $\geq \rho s$ is at most $e^{-\Omega(t)}$.*

Proof : By Lemma 3.14, G is $\left\langle\rho(1-\frac{\varepsilon}{\rho^k}), \frac{\varepsilon}{2}\right\rangle$-connected. For technical reasons, we will use the fact that this implies that G is also $\left\langle\rho(1-\frac{\varepsilon}{2\rho^k}), \frac{\varepsilon}{2}\right\rangle$-connected. Let $\rho' = \rho(1-\frac{\varepsilon}{2\rho^k})$ and $\varepsilon' = \varepsilon/2$. We would like to bound the probability that H does not have any independent sets of size ρs. Let $\delta = 1 + \frac{\varepsilon}{2\rho^k}$. Notice that $\delta\rho' < \rho$. Hence, it suffices to bound the probability that $\alpha(H) > \delta\rho' s$. This probability, in turn, is at most the probability that H has a maximum free set of size greater than $\ell = \delta\rho' s$ (Proposition 3.8).

Let $\Gamma = \ln(\delta) - \frac{\delta-1}{\delta}$. It is not hard to verify that $\Gamma = \theta((\delta-1)^2) = \theta\left(\frac{\varepsilon^2}{\rho^{2k}}\right)$ for our value of δ. By our assumption

$$s \geq ct\frac{\rho^{2k-1}}{\varepsilon^2}\log\left(\frac{\rho}{\varepsilon}\right) \geq \frac{ct}{\rho(\delta-1)^2}\left(\log(1/\rho') + \log(1+1/(\delta-1)^2)\right) \geq \frac{ct}{\rho'\Gamma}\left(\log(1/\rho') + \log(1+1/\Gamma)\right)$$

where in the above the constant c may change values from expression to expression.

Now, by Corollary 3.13, for our choice of s, the probability that H has a maximum free set of size greater than $\ell = \delta\rho' s$ is at most $s\left(\frac{1}{e^{\Gamma}}\right)^{\Omega(\delta\rho s)} \leq e^{-\Omega(t)}$. □

Roughly speaking, Theorem 1.3 states that given a $\langle\rho,\varepsilon\rangle$-connected hypergraph G, a random sample H of G of size s proportional to $2^k k!\frac{\rho^{2k}}{\varepsilon^3}$ (or larger) will not have an independent set of size ρs (with high probability). In Section 4 we continue to study the minimal value of s for which $\alpha(H) < \rho s$ with high probability, and present a lower bound of $\frac{\rho^{2k-1}}{4(k!)^2\varepsilon^2}$ on the size of s.

4 Lower bounds for the testing of $\alpha(G)$

In this section we present k-uniform hypergraphs G which are $\langle\rho,\varepsilon\rangle$-connected, but with probability $\geq 1/20$ a random sample H of G of size $s = \frac{\rho^{2k-1}}{4(k!)^2\varepsilon^2}$ is likely to have an independent set of size greater than ρs.

Lemma 4.1. *Let n be a large constant. Let $\rho > 0$ and $\varepsilon > 0$ satisfy (a) $\varepsilon << \frac{\rho^k}{2K!}$, (b) $\rho^{2k-1}/\varepsilon^2 << n$ and (c) $k^2 << \rho n$. For n large enough, there exists a graph G on n vertices for which G is $\langle\rho,\varepsilon\rangle$-connected, and with probability $\geq 1/20$ a random set H of size $s = \frac{\rho^{2k-1}}{4(k!)^2\varepsilon^2}$ will have an independent set of size ρs.*

Proof (sketch) : Consider the k-uniform hypergraph $G = (V, E)$ in which (a) $|V| = n$, (b) V consists of two disjoint sets A and $V \setminus A$, where A is of size $(1 - \frac{2k!\varepsilon}{\rho^k})\rho n$, and (c) the edge set E of G consists of all subsets of V of size k except those included in A (namely, A is an independent set). On one hand, every subset of size ρn in G induces a subgraph with at least εn^k edges (implying that G is $\langle\rho,\varepsilon\rangle$-connected). On the other, let H be a random subset of V obtained by picking each vertex independently with probability $\frac{1}{n}\frac{\rho^{2k-1}}{4(k!)^2\varepsilon^2}$. The expected size of H is $s = \frac{\rho^{2k-1}}{4(k!)^2\varepsilon^2}$. In the following, we assume H is exactly of size s, minor modifications in the proof are needed if this assumption is not made. The set $H \cap A$ is an independent set in the subgraph induced by H. The expected size of $H \cap A$ is $(1 - \frac{2k!\varepsilon}{\rho^k})\rho s$. Let $N(0,1)$ denote a standard normal variable. It can be seen using the central limit theorem (for example [Fel66]) that for our choice of parameters, the probability that $|H \cap A|$ deviates from its expectation by more than a square root of its expectation is at least

$$\Pr\left[|H \cap A| > \left(1 - \frac{2k!\varepsilon}{\rho^k}\right)\rho s + \sqrt{\rho s}\right] > \Pr\left[N(0,1) > 3/2\right] \geq \frac{1}{20}.$$

In such a case the size of $H \cap A$ will be greater than $(1 - \frac{2k!\varepsilon}{\rho^k})\rho s + \sqrt{\rho s} = \rho s$ for our value of s. Hence implying the second assertion of the lemma. □

References

[AdlVKK02] N. Alon, W. Fernandez de la Vega, R. Kannan, and M. Karpinski. Random sampling and approximation of max-csp problems. *In Proceddings of STOC*, pages 232–239, 2002.

[AS03] N. Alon and A. Shapira. Testing satisfyability. *Journal of Algorithms*, 47:87–103, 2003.

[CS01] A. Czumaj and C. Sohler. Testing hypergraph coloring. *In Proceddings of ICALP*, pages 493–505, 2001.

[Fel66] W. Feller. *An introduction to probability theory and its applications*, volume 2. John Wiley & Sons, 1966.

[Fis01] E. Fisher. The art of uninformed decisions: A primer to property testing. *The Computational Complexity Column of The Bulletin of the European Association for Theoretical Computer Science*, 75:97–126, 2001.

[FLS02] U. Feige, M. Langberg, and G. Schechtman. Graphs with tiny vector chromatic numbers and huge chromatic numbers. *In proceedings of 43rd annual Symposium on Foundations of Computer Science*, pages 283–292, 2002.

[GGR98] O. Goldreich, S. Goldwasser, and D. Ron. Property testing and its connection to learning and approximation. *Journal of ACM*, 45(4):653–750, 1998.

[Gol98] O. Goldreich. Combinatorial property testing - a survey. *Randomization Methods in Algorithm Design (P. Pardalos, S. Rajasekaran and J. Rolim eds.), AMS-DIMACS*, pages 45–60, 1998.

[Ron01] D. Ron. Property testing (a tutorial). *In Handbook of Randomized Computing (S. Rajasekaran, P. M. Pardalos, J. H. Reif and J. D. P. Rolim eds), Kluwer Press*, 2001.

[RS96] R. Rubinfeld and M. Sudan. Robust characterization of polynomials with applications to program testing. *SIAM Journal of Computing*, 25:252–271, 1996.

ECCC ISSN 1433-8092
http://www.eccc.uni-trier.de/eccc
ftp://ftp.eccc.uni-trier.de/pub/eccc
ftpmail@ftp.eccc.uni-trier.de, subject 'help eccc'