SIAM J. DISCRETE MATH.
Vol. 26, No. 3, pp. 1134–1147
© 2012 Society for Industrial and Applied Mathematics

Downloaded 02/18/13 to 131.193.153.143. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

NEW LOWER BOUNDS FOR THE INDEPENDENCE NUMBER OF SPARSE GRAPHS AND HYPERGRAPHS*

KUNAL DUTTA†, DHRUV MUBAYI‡, AND C. R. SUBRAMANIAN†

Abstract. We obtain new lower bounds for the independence number of K_r-free graphs and linear k-uniform hypergraphs in terms of the degree sequence. This answers some old questions raised by Caro and Tuza [*J. Graph Theory*, 15 (1991), pp. 99–107]. Our proof technique is an extension of a method of Caro [*New Results on the Independence Number*, Technical report, Tel Aviv University, 1979] and Wei [*A Lower Bound on the Stability Number of a Simple Graph*, TM 81-11217-9, Bell Laboratories, Berkley Heights, NJ, 1981], and we also give a new short proof of the main result of Caro and Tuza using this approach. As byproducts, we also obtain some nontrivial identities involving binomial coefficients, which may be of independent interest.

Key words. independence number, hypergraphs, random permutations

AMS subject classifications. 05D40, 05C65

DOI. 10.1137/110839023

1. Introduction.

For $k \geq 2$, a k-uniform hypergraph H is a pair $(V(H), E(H))$, where $E \subseteq \binom{V(H)}{k}$. A set $I \subset V(H)$ is an independent set of H if $e \not\subseteq I$ for every $e \in E(H)$ or equivalently, $\binom{I}{k} \cap E(H) = \emptyset$. The independence number of H, denoted by $\alpha(H)$, is the maximum size of an independent set in H. For $u \in V(H)$, its degree in H, denoted by $d_H(u)$, is defined to be $|\{e \in E(H) : u \in e\}|$. (We omit the subscript if it is obvious from the context.) Throughout this paper, we use t to denote $k-1$ except in some places where it stands for some real value. (The correct meaning can be easily inferred from the context.) Also, we use the term graph whenever k happens to be 2. A k-uniform hypergraph is *linear* if it has no 2-cycles, where a 2-cycle is a set of 2 hyperedges containing at most $2t$ vertices. The dual of the above definition says that a linear hypergraph is one in which every pair of vertices is contained in at most one hyperedge.

In [17], Turán proved a theorem giving a tight bound on the maximum number of edges that a K_r-free graph can have, which has since become the cornerstone theorem of extremal graph theory. Turán's theorem, when applied to the complement $\overline{G}$ of a graph G, yields a lower bound $\alpha(G) \geq \frac{n}{d+1}$, where d denotes the average degree in G of its vertices.

Caro [5] and Wei [18] independently proved that $\alpha(G) \geq \sum_v \frac{1}{d(v)+1}$ which is at least $\frac{n}{d+1}$. The probabilistic proof of their result later appeared in Alon and Spencer's book [4].[1] One natural extension of Turán's theorem to k-uniform hypergraphs H is the bound $\alpha(H) > c_k \frac{n}{d^{1/t}}$, and this was shown via an easy probabilistic argument by Spencer [13]. Caro and Tuza [6] improved this bound for arbitrary k-uniform

*Received by the editors June 29, 2011; accepted for publication (in revised form) June 14, 2012; published electronically August 16, 2012.
http://www.siam.org/journals/sidma/26-3/83902.html

†The Institute of Mathematical Sciences, Taramani, Chennai - 600113, India (kdutta@imsc.res.in, crs@imsc.res.in).
‡Department of Mathematics, Statistics, and Computer Science, University of Illinois, Chicago, IL 60607 (mubayi@math.uic.edu). This author's research was supported in part by NSF grant DMS 0969092.

[1]According to Bopanna [8], the probabilistic argument in [4] was obtained by him, although it is possible that it was known earlier.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

hypergraphs. In order to state their lower bound, we need the following definition (of fractional binomial coefficients) from [9].

DEFINITION. For $t > 0$, $a \geq 0$, $d \in \mathbb{N}$

$$\binom{d+1/t}{a} := \frac{(td+1)(t(d-1)+1)\dots(t(d-a+1)+1)}{a!t^a}.$$

What Caro and Tuza [6] showed was that

$$\alpha(H) \geq \sum_{v \in V(H)} \frac{1}{\binom{d(v)+1/t}{d(v)}}. \tag{1}$$

Indeed, an easy consequence of (1) is the following result.

THEOREM 1.1 (see Caro–Tuza [6]). *For every $k \geq 3$, there exists $d_k > 0$ such that every k-uniform hypergraph H has*

$$\alpha(H) \geq d_k \sum_{v \in V(H)} \frac{1}{(d(v)+1)^{1/t}}.$$

As a corollary, one infers the bound of Spencer above. Later, Thiele [16] provided a lower bound on the independence number of nonuniform hypergraphs, based on the degree rank (a generalization of degree sequence).

In this paper, we prove new lower bounds for the independence number of locally sparse graphs and linear k-uniform hypergraphs. The starting point of our approach is the probabilistic proof of Boppana–Caro–Wei. This approach, together with some additional simple ideas, quickly yields a new short proof of Theorem 1.1. (See section 2 for the detailed proof.)

1.1. K_r-free graphs. For certain classes of sparse graphs, improvements of the Caro–Wei bound (in terms of average degree d) are known. Ajtai, Komlós, and Szemerédi [2] proved a lower bound of $\Omega(\frac{n \log d}{d})$ for the independence number of triangle-free graphs. An elegant and simpler proof was later given by Shearer [10], who also improved the constant involved. Ajtai et al. [1] showed that for K_r-free graphs ($r > 3$), the independence number is lower-bounded by $c_r(n/d)\log(\frac{(\log d)}{r})$, where $c_r \in \Re^+$ depends only on r. They also conjectured that the optimal bound is $c_r \frac{n \log d}{d}$. Shearer [12] improved their bound to $\Omega(\frac{n \log d}{d \log \log d})$.

Caro and Tuza [6] raised the following question in their 1991 paper.

(i) Can the lower bounds of Ajtai et al. [2] and Shearer [10, 12] be generalized in terms of degree sequences?

We answer this question via the following two theorems.

THEOREM 1.2. *For every $\epsilon \in (0,1)$ there exists $c > 0$ such that the following holds: Every triangle-free graph G with average degree D has independence number at least*

$$c(\log D) \sum_{v \in V(G)} \frac{1}{\max\{D^\epsilon, d(v)\}}.$$

THEOREM 1.3. *For every $\epsilon \in (0,1)$ and $r \geq 4$, there exists $c > 0$ such that the following holds: Every K_r-free graph G with average degree D has independence number at least*

$$c \frac{\log D}{\log \log D} \sum_{v \in V(G)} \frac{1}{\max\{D^\epsilon, d(v)\}}.$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/18/13 to 131.193.153.143. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 02/18/13 to 131.193.153.143. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

A similar bound to Theorem 1.2 was obtained by Shearer [11], who showed that for a triangle-free graph G, $\alpha(G) \geq (1-o(1))\sum_{v\in V(G)} \frac{\log d(v)}{d(v)}$. We remark that up to multiplicative constants, the function $\log D \sum_v \frac{1}{d(v)}$ is larger than the function $\sum_v \frac{\log d(v)}{d(v)}$ used in Shearer's bound.

1.2. Linear hypergraphs. As mentioned earlier, a lower bound of $\Omega(n/d^{1/t})$ for an n vertex k-uniform hypergraph with average degree d can be inferred from Theorem 1.1. Caro and Tuza [6] also raised the following question.

(ii) How can one extend the lower bounds of Ajtai et al. [2] and Shearer [10, 12] to hypergraphs?

As it turns out, such extensions were known for the class of linear k-uniform hypergraphs. Indeed, the lower bound

$$\alpha(H) = \Omega\left(n\left(\frac{\log d}{d}\right)^{1/t}\right), \tag{2}$$

where H is a linear k-uniform hypergraph with average degree d was proved by Duke, Lefmann, and Rödl [7] using the results of [3]. Our final result generalizes (2) in terms of the degree sequence of the hypergraph.

THEOREM 1.4. *For every $k \geq 3$ and $\epsilon \in (0,1)$, there exists $c > 0$ such that the following holds: Every linear k-uniform hypergraph H with average degree D has independence number at least*

$$c(\log D)^{1/t} \sum_{v\in V(H)} \frac{1}{\max\{D^{\epsilon/t}, (d(v))^{1/t}\}}.$$

We also describe an infinite family of k-uniform linear hypergraphs to illustrate that the ratio between the bounds of Theorem 1.4 and (2) can be unbounded in terms of the number of vertices.

The remainder of this paper is organized as follows. In section 2, we give a new short proof of Theorem 1.1. In section 3, we apply the analysis in section 2 to the special case of linear hypergraphs and obtain a "warm-up" result, Theorem 3.1, which will be helpful in proving the main technical result, Theorem 4.1, proved in section 4. The expression obtained in Theorem 4.1 plays a crucial role in the proofs of Theorems 1.2, 1.3, and 1.4; these are provided in section 5. In section 6, we give infinite families of K_r-free graphs and k-uniform linear hypergraphs which illustrate that the bounds in Theorems 1.2, 1.3, and 1.4 can be bigger than the corresponding bounds in [2, 3, 7, 10, 12] by arbitrarily large multiplicative factors. Finally, in section 8, we state several combinatorial identities which follow as simple corollaries of Theorem 4.1.

2. A new proof of Theorem 1.1.

In this section we obtain a new short proof of Theorem 1.1. First we obtain the following theorem which is later used to prove Theorem 1.1.

THEOREM 2.1. *For every $k \geq 2$, there exists a constant $c = c_k$ such that any k-uniform hypergraph H on n vertices and $m \geq 1$ hyperedges satisfies*

$$\sum_{J\subset V(H)} \frac{1}{\binom{n}{|J|}} > c\frac{n}{m^{1/k}}, \tag{A}$$

where we sum over all independent sets J.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/18/13 to 131.193.153.143. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Proof. Let $t_k(n,m)$ denote the left-hand side (LHS) of (A). Consider any edge $e \in E(H)$. The edge e can belong to at most $\binom{n-k}{j-k}$ nonindependent sets of size j. Since there are m edges there are at most $m\binom{n-k}{j-k}$ sets of size j that are not independent. Thus, at least $\binom{n}{j} - m\binom{n-k}{j-k}$ sets of size j are independent. Hence we have

$$\begin{aligned} t_k(n,m) &\geq \sum_{j=1}^{n} \left(1 - m\frac{\binom{n-k}{j-k}}{\binom{n}{j}}\right) \quad = \quad \sum_{j=1}^{n} \left(1 - m\frac{(j)_k}{(n)_k}\right) \\ &> \sum_{j=1}^{\lfloor n/(2m)^{1/k} \rfloor} \left(1 - m\frac{j^k}{n^k}\right) \quad \geq \quad \sum_{j=1}^{\lfloor n/(2m)^{1/k} \rfloor} \left(1 - m\frac{1}{2m}\right) \\ &\geq \frac{1}{2}\left\lfloor \frac{n}{(2m)^{1/k}} \right\rfloor \quad \geq \quad c_k \frac{n}{m^{1/k}} \end{aligned}$$

for some suitably chosen c_k which is close to $2^{-(k+1)/k}$. ∎

Let $H = (V,E)$ be a k-uniform hypergraph. For $k \geq 3$ and for $u \in V$ with $d_H(u) \geq 1$, the link graph associated with u in H is the t-uniform hypergraph $L_u = (U,F)$, where $U := \{v \neq u : \exists e \in E : \{u,v\} \subseteq e\}$ and $F = \{e \setminus u : u \in e \in E\}$. Let $\mathcal{I}(H)$ denote the collection of independent sets of H.

Proof of Theorem 1.1. As mentioned in the introduction, the proof is an extension of the technique used in Alon and Spencer's book [4]. Let $H = (V,E)$ be an arbitrary k-uniform hypergraph. Choose uniformly at random a total ordering $<$ on V. Define an edge $e \in E$ to be *backward* for a vertex $v \in e$ if $u < v$ for every $u \in e \setminus \{v\}$. Define a random subset I to be the set of those vertices v such that no edge e incident at v is backward for v with respect to $<$. Clearly, I is independent in H. We have $E[|I|] = \sum_v Pr(v \in I)$. If $d_v = 0$, then $v \in I$ with probability 1. Hence, we assume that $d(v) \geq 1$. From the definition of I, it follows that $v \in I$ if and only if for every e incident at v, $e \setminus \{v\} \not\subseteq S_v \; = \{u \in V(L_v) : u < v\}$. In other words, S_v is an independent set in L_v. Let $l_v = |V(L_v)|$. Then

$$Pr[v \in I] = \sum_{J \in \mathcal{I}(L_v)} \frac{|J|!(l_v - |J|)!}{(l_v+1)!} \quad = \quad \frac{1}{l_v+1} \sum_{J \in \mathcal{I}(L_v)} \frac{1}{\binom{l_v}{|J|}}.$$

Applying Theorem 2.1 to the t-uniform link graph L_v (with $c = c_{k-1}$), we get

$$Pr[v \in I] \geq \frac{c}{l_v+1}\left(\frac{l_v}{d(v)^{1/(k-1)}}\right) \quad \geq \quad \frac{cl_v}{l_v+1}\left(\frac{1}{(d(v)+1)^{1/(k-1)}}\right).$$

Since $l_v \geq k-1$, we get $Pr[v \in I] \geq ((k-1)c/k)\frac{1}{(d(v)+1)^{1/(k-1)}}$. By choosing $d_k = (k-1)c/k$, we get the lower bound of the theorem. ∎

3. Linearity: Probability of having no backward edges. In this section, we state and prove a warm-up result on the probability of having no backward edges incident at a vertex for a randomly chosen linear ordering (Theorem 3.1 below). The problem is the same as in the previous section, only now the hypergraph under consideration is assumed to be linear and we get an explicit closed-form expression for this probability. This result will be helpful for the proof of the main technical theorem, given in the next section.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/18/13 to 131.193.153.143. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

THEOREM 3.1. *Let H be a linear k-uniform hypergraph and let v be an arbitrary vertex having degree d. For a uniformly chosen total ordering $<$ on V, the probability $P_v(0)$ that v has no backward edge incident at it is given by*

$$P_v(0) = \frac{1}{\binom{d+1/t}{d}}.$$

Remark. It is interesting to note that the above expression when summed over all vertices is the same bound which Caro and Tuza obtain in [6] (using very different methods), although their bound holds for independent sets in *general* k-uniform hypergraphs.

We prove the theorem using the well-known principle of inclusion and exclusion (PIE). First we state an identity involving binomial coefficients.

LEMMA 3.2. *Given nonnegative integers d and t,*

$$\sum_{r=0}^{d}(-1)^r\binom{d}{r}\frac{1}{tr+1} = \frac{1}{\binom{d+1/t}{d}}.$$

For proof see [9, equation 5.41]. Alternatively, it can be proved using the Chu–Vandermonde identity (see, e.g., [9, equation 5.93]).

Proof of Theorem 3.1. First, observe that since H is linear, the number of vertices that are neighbors of v is exactly $(k-1)d = td$. Next, notice that since the random ordering is uniformly chosen, only the relative arrangement of these td neighbors and the vertex v, i.e., $td+1$ vertices in all, will determine the required probability. Hence the total number of orderings under consideration is $(td+1)!$.

Label the hyperedges incident at v with $1,\dots,d$ arbitrarily. For a permutation π, we say that π has the property $T_{\geq S}$ if the edges with labels in S, $S \subseteq [d]$ are backward. Also, we say π has the property $T_{=S}$ if the edges with labels in S are backward and no other edges are backward. For a set S of hyperedges incident at v, let $N(T_{\geq S})$ denote the number of orderings having the property $T_{\geq S}$, that is, the number of permutations such that the hyperedges in S will *all* be backward edges. $N(T_{=S})$ is similarly defined. $N(T_{\geq S})$ is determined as follows.

Suppose S has r hyperedges incident at v. For a *fixed* arrangement of the vertices belonging to edges in S, the number of permutations of the remaining vertices is $(td+1)!/(tr+1)!$. In each allowed permutation, the vertex v must occur only after the vertices of S (i.e., the rightmost position). However the remaining tr vertices can be arranged among themselves in $(tr)!$ ways. Thus we have

$$N(T_{\geq S}) = (td+1)!\frac{(tr)!}{(tr+1)!} = \frac{(td+1)!}{(tr+1)}.$$

Clearly, if a permutation has the property $T_{\geq S}$, it has the property $T_{=S'}$ for some $S' \supseteq S$. Hence for every $S \subset [d]$,

$$N(T_{\geq S}) = \sum_{S' \supseteq S} N(T_{=S'}).$$

Therefore, by PIE (see [14, Chapter 2]),

$$N(T_{=\emptyset}) = \sum_{S}(-1)^{|S|}N(T_{\geq S}),$$

$$\sum_{|S|=r} N(T_{\geq S}) = \binom{d}{r}N(T_{\geq [r]}) \quad = \quad \binom{d}{r}\frac{(td+1)!}{tr+1}.$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Hence we get the required probability to be

$$\begin{aligned} P_v(0) &= \left(\sum_{r=0}^{d} \binom{d}{r} (-1)^r \frac{(td+1)!}{tr+1} \right) \times \frac{1}{(td+1)!} \\ &= \sum_{r=0}^{d} \binom{d}{r} (-1)^r \frac{1}{tr+1}. \end{aligned}$$

By Lemma 3.2,

$$P_v(0) = \frac{1}{\binom{d+1/t}{d}},$$

and this completes the proof. ∎

4. Linearity: Probability of having few backward edges. Now, we consider the more general case when at most $A-1$ backward edges are allowed. In this section, we get an exact expression for the corresponding probability. This estimate plays an important role later in getting new and improved lower bounds on $\alpha(H)$ for locally sparse graphs and linear hypergraphs. Our goal in this section is to prove the following result.

THEOREM 4.1. *For a k-uniform linear hypergraph H with a vertex v having degree d, a uniformly chosen permutation π induces at most $A-1$ backward edges with probability $P_v(A-1)$ given by*

$$P_v(A-1) = \begin{cases} 1 & \text{if } d \le A-1, \\ \frac{tA}{tA+1} \left[\binom{d}{A} / \binom{d+1/t}{d-A} \right] & \text{if } d \ge A. \end{cases}$$

COROLLARY 4.2. *As $d \to \infty$, the asymptotic expression for the probability $P_v(A-1)$ is given by*

$$P_v(A-1) \ \sim \ \frac{1}{1+(1/(tA))} \left(\frac{A}{d} \right)^{1/t} \ = \ \Omega((A/d)^{1/t}).$$

Proof. The asymptotics are w.r.t. $d \to \infty$, $d \ge A$. The expression for having at most $A-1$ backward edges is

$$\begin{aligned} P_v(A-1) &= \frac{1}{1+(tA)^{-1}} \frac{d(d-1)\dots(A+1)}{(d-A)!} \frac{(d-A)!}{(d+1/t)(d-1+1/t)\dots(A+1+1/t)} \\ &= \frac{1}{1+(tA)^{-1}} \frac{1}{(1+1/td)(1+(t(d-1))^{-1})\dots(1+(t(A+1))^{-1})}. \end{aligned}$$

Now, for $0 < x$ we have $(1+x)^{-1} > e^{-x}$. So we get

$$\begin{aligned} P_v(A-1) &> (1+(tA)^{-1})^{-1} e^{(-1/t)\sum_{r=A+1}^{d}(1/r)} \\ &= (1+(tA)^{-1})^{-1} e^{(-1/t)[\sum_{r=1}^{d}(1/r) - \sum_{r=1}^{A}(1/r)]} \\ &= (1+(tA)^{-1})^{-1} e^{(-1/t)[\ln d - \ln A] + O((d-A)/(tdA))} \\ &= (1+(tA)^{-1})^{-1} e^{(-1/t)\ln(d/A) - O((d-A)/(tdA))} \\ &= (1+(tA)^{-1})^{-1} (A/d)^{1/t} \Omega(1) \\ &= \Omega((A/d)^{1/t}). \end{aligned}$$

The above expression therefore becomes $\Omega((A/d)^{1/t})$. ∎

Downloaded 02/18/13 to 131.193.153.143. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/18/13 to 131.193.153.143. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

The version of PIE used most commonly deals with $N(T_{=\emptyset})$, i.e., the number of elements in the set of interest—in this case, permutations of $[td+1]$ which do not have *any* of the properties under consideration (in this case, backward edges with respect to v). However we need something slightly different—an expression for the number of permutations which have *at least* A backward edges. Clearly, the remaining permutations are those which have *at most* $A-1$ backward edges.

Therefore, we use a slightly modified version of PIE, which is stated below in Theorem 4.5. This form is well known (see, e.g., [14, Chapter 2, Exercise 1]), although it seems to be used less frequently. For the sake of completeness, we provide a simple proof. First we state two identities involving binomial coefficients. The first can be proved easily by induction on b and we omit the proof, and the second is proved in the appendix.

LEMMA 4.3. *For a, b nonnegative integers,*

$$\sum_{i=0}^{b}(-1)^i\binom{a+b}{a+i}\binom{a+i-1}{i} \;=\; 1.$$

LEMMA 4.4. *Given nonnegative integers d, A, $d \geq A$ and a positive integer t,*

$$\sum_{r=0}^{d-A}(-1)^r\binom{d}{r+A}\binom{A+r-1}{r}\frac{1}{t(r+A)+1} \;=\; 1-\left(\frac{At}{tA+1}\right)\frac{\binom{d}{A}}{\binom{d+1/t}{d-A}}.$$

We now present the generalized PIE and its well-known proof.

THEOREM 4.5. *Let S be an n-set and $E_1, E_2, \dots, E_d$ not necessarily distinct subsets of S. For any subset M of $[d]$, define $N(M)$ to be the number of elements of S in $\cap_{i\in M}E_i$ and for $0 \leq j \leq d$, define $N_j := \sum_{|M|=j} N(M)$. Then the number $N_{\geq a}$ of elements of S in* at least *a, $0 \leq a \leq d$ of the sets E_i, $1 \leq i \leq d$, is*

$$N_{\geq a} = \sum_{i=0}^{d-a}(-1)^i\binom{a+i-1}{i}N_{i+a}. \tag{MPIE}$$

Proof. Take an element $e \in S$.

(i) Suppose e is in no intersection of at least a E_i's. Then e does not contribute to any of the summands in the right-hand side (RHS) of the expression (MPIE), and hence, its net contribution to the RHS is zero.

(ii) Suppose e belongs to exactly $a+j$ of the E_i's, $0 \leq j \leq d-a$. Then its contribution to the RHS of (MPIE) is

$$\sum_{l=0}^{j}(-1)^l\binom{a+j}{a+l}\binom{a+l-1}{l},$$

and by Lemma 4.3 this is equal to 1. ∎

Proof of Theorem 4.1. If $d \leq A-1$, then $P_v(A-1) = 1$ obviously. The proof is similar to the proof of Theorem 3.1, except that in place of the PIE, we use Theorem 4.5. The set under consideration is the set of permutations of $[td+1]$; the subsets E_i correspond to the permutations for which the ith edge is backward. It is easy to see that $N(M) = N(T_{\geq M})$ under the notation used in Theorem 3.1 and hence $N(M) = \frac{(td+1)!}{t|M|+1}$. Therefore we have $N_j = \binom{d}{j}\frac{(td+1)!}{tj+1}$ as before. Hence the expression for the probability $Q_v(A)$ that *at least* A edges are backward under a uniformly

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

random permutation π becomes

$$Q_v(A) = \sum_{i=0}^{d-A} (-1)^i \binom{d}{i+A} \binom{A+i-1}{i} \frac{1}{t(i+A)+1}.$$

By Lemma 4.4 the RHS of the above expression is

$$Q_v(A) \quad = \quad 1 - \left(\frac{1}{1+(tA)^{-1}} \right) \frac{\binom{d}{A}}{\binom{d+1/t}{d-A}}.$$

Hence the probability of having at most $A-1$ backward edges is given by

$$P_v(A-1) \quad = \quad \frac{1}{1+(tA)^{-1}} \frac{\binom{d}{A}}{\binom{d+1/t}{d-A}}$$

and the proof is complete. $\square$

5. Lower bounds for linear hypergraphs and K_r-free graphs. In this section we prove Theorems 1.2, 1.3, and 1.4. These follow by a simple application of Corollary 4.2. Since the proofs follow the same outline, we prove them simultaneously, highlighting only the differences as and when they occur.

Proofs of Theorems 1.2, 1.3, *and* 1.4. Consider a uniformly chosen random permutation of the vertices of the graph/hypergraph under consideration. Let D be the average degree of the graph or hypergraph and $A = D^\epsilon$. Let I be the set of those vertices each having at most $A-1$ backward edges incident on it. Clearly, the expected size of I is

$$\begin{aligned} E[|I|] = \sum_{v \in V} P_v(A-1) \quad &\geq \quad c \sum_{v \in V} \left(\frac{A}{\max\{A, d(v)\}} \right)^{1/t} \\ = cA^{1/t} \sum_{v \in V} &\left(\frac{1}{\max\{A, d(v)\}} \right)^{1/t} \end{aligned}$$

for some constant $c = c(k, \epsilon)$. (For a graph, $k = 2$ and hence $t = 1$.) Also, by construction, the average degree of the sub(hyper)graph induced by I is at most $k(A-1)$. Therefore, there exists an independent set I' of size at least as follows:

(i) Case $t = 1$, graph is K_3-free: By [10], $\alpha(G)$ is at least

$$\Omega\left(\log(2(A-1)) \frac{|I|}{2(A-1)} \right) \quad = \quad \Omega\left(\log D \sum_{v \in V} \frac{1}{\max\{A, d(v)\}} \right).$$

(ii) Case $t = 1$, graph is K_r-free ($r > 3$): By [12], $\alpha(G)$ is at least

$$\Omega\left(\frac{\log(2(A-1))}{\log\log(2(A-1))} \frac{|I|}{2(A-1)} \right) \quad = \quad \Omega\left(\frac{\log D}{\log\log D} \sum_{v \in V} \frac{1}{\max\{A, d(v)\}} \right).$$

(iii) Case $t > 1$, hypergraph is linear: By [7], $\alpha(H)$ is at least

$$\Omega\left((\log k(A-1))^{1/t} \frac{|I|}{(k(A-1))^{1/t}} \right) \quad = \quad \Omega\left((\log D)^{1/t} \sum_{v \in V} \frac{1}{(\max\{A, d(v)\})^{1/t}} \right).$$

The above three cases prove Theorems 1.2, 1.3, and 1.4, respectively.

Downloaded 02/18/13 to 131.193.153.143. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/18/13 to 131.193.153.143. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Note. An inspection of the proofs above show why we need ϵ to be a fixed constant. It is because all three expressions above essentially have $\log A$, i.e., $\epsilon \log D$ in the numerator. So, if $\epsilon = o(1)$, then $\log A = o(\log D)$, and we would get asymptotically weaker results. ∎

6. Construction comparing average-degree versus degree-sequence-based bounds. A degree-sequence-based bound obviously reduces to a bound based on average degree, when the (hyper)graph is regular. However, the convexity of the function $x^{-1/t}$, $x \geq 1$ and $t \in \mathbb{N}$ shows that the bounds in Theorems 1.2, 1.3, and 1.4 are better than the corresponding average-degree-based bounds proved in [3, 10, 12], respectively, provided the minimum degree is at least A, although it is not clear a priori if the improvement can become significantly larger. Also, at least half the vertices will have degree at most $2D$, so even in the general case (no restriction on the minimum degree) our bounds are no worse than the average-degree-based bounds (ignoring the constant factors). In fact, they can be much larger than the latter bounds. We now give infinite families of K_r-free graphs and linear k-uniform hypergraphs which show

(i) The bounds given by Theorems 1.2 and 1.3 can be better than the bounds in [2, 10, 12], respectively, by a multiplicative factor of $\log(|V(G)|)$.

(ii) The bound in Theorem 1.4 can be better than the bound in [3] by a multiplicative factor of $((\log|V(H)|)/(\log\log|V(H)|))^{(1-\epsilon)/t}$, where ϵ is the constant mentioned in Theorem 1.4.

Case (i). Take a set of n disjoint graphs, $K_{1,1}$, $K_{2,2}$, $K_{4,4}, \ldots, K_{2^{n-1},2^{n-1}}$. For each $i \in [n]$, join one of the parts of the component $K_{2^i,2^i}$ to one of the parts in $K_{2^{i-1},2^{i-1}}$ by introducing a complete bipartite graph between them. (Use the other part of $K_{2^i,2^i}$ for joining to $K_{2^{i+1},2^{i+1}}$). Let G denote the resulting connected triangle-free graph.

The total number of vertices is $2^{n+1}-2$, whereas the average degree is $d_{av} = 2|E(G)|/|V(G)| = (2^n+1)/2 - o(1)$. Hence, the average-degree-based bound gives $\Theta(|V(G)|\log d_{av}/d_{av}) = \Theta(\log d_{av})$. Denote by l the maximum j such that $3.2^j \leq A < 3.2^{j+1}$, where $A := d_{av}^{\epsilon}$. For every fixed $\epsilon \in (0,1)$, we have $n - l = \Theta(n)$. Theorem 1.2 gives

$$\begin{aligned} c\log d_{av} \sum_{v\in V} \frac{1}{\max\{d(v),A\}} &= c\log d_{av}\left[\frac{1}{A} + \sum_{j=1}^{l}\frac{3.2^j}{A} + \sum_{j=l+1}^{n-2}\left(\frac{3.2^j}{3.2^j}\right) + \frac{2^{n-1}}{2^{n-1}}\right] \\ &= c(\log d_{av})\left[\Theta(1)+\Theta(n)\right] \\ &= c(\log d_{av})\Theta(\log(|V(G)|). \end{aligned}$$

The same example works for Theorem 1.3 also, since triangle-free graphs are obviously K_r-free for $r \geq 3$.

Case (ii). Fix some $m = m(n) = k^{2^n}$. For each $i \in \{0,\ldots,n-1\}$, first create a connected linear hypergraph as follows: Take the vertex set as $[k]^{2^i}$, i.e. the set of 2^i-dimensional vectors with each coordinate of a vector taking values in $\{1,2,\ldots,k\}$. Let each hyperedge consist of the k vertices which have all but one coordinate fixed. Call this hypergraph an *i-unit*. It can be verified easily that each i-unit is k-uniform and 2^i-regular. Now for each i, create an *i-component* as follows:

(i) Take $m_i = \lceil \frac{m}{k^{2^i}} \rceil$ disjoint unions of i-units and linearly order them, say $i_1, \ldots, i_{m_i}$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/18/13 to 131.193.153.143. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

(ii) Consider the sets of vertices of size k formed by choosing at most one vertex from each i-unit. Add such edges greedily, ensuring the following:
 (i) No vertex belongs to more than one such edge;
 (ii) The first edge is chosen from i-units $i_1, \ldots, i_k$, the second one from $i_2, \ldots, i_{k+1}$, etc. In general the jth such edge has one vertex from each of the i-units $i_{j \pmod{m_i}}, i_{j+1 \pmod{m_i}}, \ldots i_{j+k-1 \pmod{m_i}}$.

Each i-component is a connected, linear k-uniform graph and the degree of every vertex is either 2^i or $2^i + 1$. Take the disjoint unions of n such i-components, one for every $i \in \{0, \ldots, n-1\}$, to get the hypergraph $H = H(n) = (V, E)$. For each $j \in \{0, \ldots, n-2\}$, greedily add a maximal matching between components j and $j+1$ with each edge taking only one vertex from component j (and remaining $k-1$ from the component $j+1$) and no vertex belonging to more than one such edge. Let G be the resulting connected, linear k-uniform graph. The total number of vertices in the jth component is $m_j \cdot k^{2^j} = m(1 + o(1))$, and hence $|V| = nm(1 + o(1))$. Also, the average degree is $d_{av} \sim (2^n - 1)/n \sim 2^n/n$. Let l denote the greatest integer j such that $2^j \leq (d_{av})^\epsilon \sim 2^{\epsilon n}/n^\epsilon$. Therefore the average-degree-based bounds in [3, 7] give a lower bound of

$$\alpha(H) = \Omega(mn^{1+1/t}(\log d_{av})^{1/t}/2^{n/t}). \tag{A}$$

Notice that the degree of any vertex in the ith component (after G has been constructed) is always between 2^i and $2^i + 3$. For a vertex v such that $d(v) < d_{av}^\epsilon$, the actual degree does not play a role in the expression in Theorem 1.4. For vertices v such that $d(v) \geq d_{av}^\epsilon$, this increase is negligible $(2^{\epsilon n}/n^\epsilon + 3) \sim 2^{\epsilon n}/n^\epsilon$. Therefore, the bound in Theorem 1.4 gives

$$\begin{aligned}
\alpha(H) &= \Omega\left((\log d_{av})^{1/t}\left[\sum_{j=0}^{l} \frac{mn^{\epsilon/t}}{2^{\epsilon n/t}} + \sum_{j=l+1}^{n-1} \frac{m}{2^{j/t}}\right]\right)\\
&= \Omega\left(m(\log d_{av})^{1/t}\left[\epsilon 2^{-\epsilon n/t} n^{1+\epsilon/t} + 2^{-\epsilon n/t} n^{\epsilon/t} \frac{(1 - 2^{-(n-l-1)/t})}{1 - 2^{-1/t}}\right]\right)\\
&= \Omega\left(m(\log d_{av})^{1/t} \times 2^{-\epsilon n/t}\left[\epsilon n^{1+\epsilon/t} + n^{\epsilon/t} \frac{(1 - 2^{-(n-l-1)/t})}{1 - 2^{-1/t}}\right]\right)\\
&= \Omega\left(m(\log d_{av})^{1/t} \times 2^{-\epsilon n/t}(\epsilon n^{1+\epsilon/t} + \Theta(n^{\epsilon/t}))\right).
\end{aligned} \tag{B}$$

The ratio of the bound in (B) to the one in (A) can be seen to be $\Omega((2^n/n)^{(1-\epsilon)/t})$, which is $\Omega((\log |V|/\log\log |V|))^{(1-\epsilon)/t})$.

7. Binomial identities. In the course of this paper, certain nontrivial binomial identities were also obtained with semicombinatorial proofs. Some of the identities are new, to the best of our knowledge, and may be of independent interest. These are described below:

$$\sum_{a=0}^{A} \sum_{i=0}^{d-a} \binom{d}{a+i} \binom{a+i}{i} 2^i (2d - 2a - i)!(2a + i)! = (d!)^2 4^{d-A}(A + 1)\binom{2A + 1}{A}. \tag{3}$$

The LHS (when divided by $(2d+1)!$) amounts to the expression for $P_v(A)$ when $k = 3$: choose $a + i$ hyperedges from the d hyperedges incident on v; of these a hyperedges are backward, while i hyperedges each have one vertex occurring prior to

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/18/13 to 131.193.153.143. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

v in the random permutation. These i vertices can be chosen from i pairs in 2^i ways. The $(2a+i)$ vertices before v can be arranged in $(2a+i)!$ ways among themselves. The remaining $(2d-2a-i)$ vertices occur after v and can be arranged among themselves in $(2d-2a-i)!$ ways. The RHS is easily obtained from Theorem 4.1 by taking $t=2$.

Even the $A=0$ case of the above identity gives (after some rearrangements)

$$\sum_{i=0}^{d} \binom{d+i}{d} 2^{-i} \quad = \quad 2^d.$$

The above identity merits discussion in some detail in [9, Chapter 5, equations 5.20, 5.135-8]; a nice combinatorial proof of it is provided in [15].

The next identity (for the more general case $k \geq 3$) is much more complicated. Given $i \in \mathbb{Z}^+$, let C_i^{t-1} denote the set of all solutions in nonnegative integers $\mathbf{j} = (j_1, j_2, \ldots, j_{t-1})$ of the equation $\mathbf{j} \cdot \mathbf{1} = i$, i.e., $C_i^{t-1} := \{(j_1, \ldots, j_{t-1}) : \sum_{l=1}^{t-1} j_l = i\ ;\ \forall l\ j_l \geq 0\}$. Then

$$\begin{aligned} &\sum_{a=0}^{A}\sum_{i=0}^{d-a}\sum_{\mathbf{j}\in C_i^{t-1}} \left[\binom{d}{a+i}\binom{a+i}{a, j_1, \ldots, j_{t-1}} \left(ta + \sum_{s=1}^{t-1} s.j_s\right)!,\right. \\ &\qquad\qquad \left.\left(td - \left(ta + \sum_{s=1}^{t-1} s.j_s\right)\right)! \prod_{r=1}^{t-1} \binom{t}{r}^{j_r}\right] \\ &= (td+1)!(1+(tA+t)^{-1})^{-1} \left[\binom{d}{A+1} \Big/ \binom{d+1/t}{d-A-1}\right]. \end{aligned} \tag{4}$$

The LHS again follows by similar arguments as for (3), this time for general t. There are a backward edges, i_1 edges which have one vertex before v, i_2 edges with 2 vertices before v, and so on. The RHS follows from Theorem 4.1.

Our proof techniques for identities (3, 4) involving PIE are nonstandard. It may be an interesting problem in enumerative combinatorics to come up with combinatorial proofs of the identities (3, 4). In particular, for (4) it would be interesting to come up with proofs using *any* standard technique such as induction, generating functions, the Wilf–Zeilberger method, etc.

8. Concluding remarks.

As the constructions of section 6 show, our degree-sequence-based lower bounds can be asymptotically better than the previous average-degree-based bounds. This is in spite of using the previous bounds in the proof. The power of the random permutation method is that it allows us to obtain a relatively large *sparse* induced subgraph, over which the application of the average-degree bound yields a much better result than a straightforward application over the entire graph would have.

With regard to the tightness of our results and the weakening parameter A, first, from the proof of Theorems 1.2–1.4, it is clear that $\epsilon = \log A/\log D$ has to be at least a constant. Ideally, we may want to have $\epsilon = 0$ in the bounds of Theorems 1.2, 1.3, and 1.4. The following example, however, shows that it is possible to construct a triangle-free graph for which the bound in, say, Theorem 1.2 would give a value more than the number of vertices: Take a disjoint union of $A = K_{n/3,n/3}$ and $B = \overline{K}_{n/3}$ and introduce a perfect matching between B and one of the parts of A. Now, $|V| = n$, $D \sim 2n/9$, and hence if $\epsilon = 0$, Theorem 1.2 would give a lower bound of $\Omega(n \log n)$, which is asymptotically larger than $|V|$. Similar examples can be constructed with linear hypergraphs also.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/18/13 to 131.193.153.143. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

9. Appendix.

Proof of Lemma 4.4. Let the LHS be denoted by S_d. Then, using the identity

$$\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1},$$

we have

$$\begin{aligned} S_d &= \sum_{r=0}^{d-A} (-1)^r \left[\binom{d-1}{r+A} + \binom{d-1}{r+A-1} \right] \binom{A+r-1}{r} \frac{1}{tr+tA+1} \\ &= S_{d-1} + \sum_{r=0}^{d-A} (-1)^r \binom{d-1}{r+A-1} \binom{A+r-1}{r} \frac{1}{tr+tA+1} \end{aligned}$$

since $\binom{d-1}{d} = 0$. Now the second sum can be simplified as

$$\begin{aligned} T_d &= \sum_{r=0}^{d-A} (-1)^r \binom{d-1}{r+A-1} \binom{A+r-1}{r} \frac{1}{tr+tA+1} \\ &= \left(\frac{(d-1)!}{(d-A)!(A-1)!} \right) \sum_{r=0}^{d-A} (-1)^r \binom{d-A}{r} \frac{1}{tr+tA+1} \\ &= \binom{d-1}{A-1} \frac{1}{tA+1} \sum_{r=0}^{d-A} (-1)^r \binom{d-A}{r} \frac{1}{(t/(tA+1))r+1}. \end{aligned}$$

By Lemma 3.2, we get

$$T_d = \frac{1}{tA+1} \left[\binom{d-1}{A-1} \Big/ \binom{d-A+(tA+1)/t}{d-A} \right].$$

Therefore,

$$S_d = S_{d-1} + \frac{1}{tA+1} \left[\binom{d-1}{A-1} \Big/ \binom{d+1/t}{d-A} \right].$$

Unraveling the recursion and noticing that $S_A = 1/(tA+1)$, we get that

$$\begin{aligned} S_d &= (1/(tA+1)) \sum_{r=0}^{d-A} \left[\binom{d-r-1}{A-1} \Big/ \binom{d+1/t-r}{d-A-r} \right] \\ &= (1/(tA+1)) \sum_{r=0}^{d-A} \left[\binom{A-1+r}{A-1} \Big/ \binom{A+1/t+r}{r} \right] \end{aligned}$$

by reversing the order of summation. Finally, the following claim completes the proof.

Claim. For $d \geq A, t \geq 0$,

$$\frac{1}{tA+1} \sum_{r=0}^{d-A} \frac{\binom{A-1+r}{A-1}}{\binom{A+1/t+r}{r}} \quad = \quad 1 - \frac{tA}{tA+1} \frac{\binom{d}{A}}{\binom{d+1/t}{d-A}}.$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Proof of Claim. We use induction on d. When $d = A$, the LHS is $(tA+1)^{-1}$, while the RHS is $1 - \frac{At}{tA+1}$, so we have equality. Now assume equality for d and consider the LHS for $d+1$:

$$\begin{aligned}
&\frac{1}{tA+1}\sum_{r=0}^{d-A+1}\frac{\binom{A-1+r}{r}}{\binom{A+1/t+r}{r}}\\
&= 1 - \frac{At}{tA+1}\left[\binom{d}{A}\Big/\binom{d+1/t}{d-A}\right] + (At+1)^{-1}\left[\binom{d}{d-A+1}\Big/\binom{d+1+1/t}{d-A+1}\right]\\
&= 1 - \frac{At}{(tA+1)\binom{d+1+1/t}{d-A+1}}\left[\binom{d}{A}\frac{d+1+1/t}{d-A+1} - (At)^{-1}\binom{d}{d-A+1}\right]\\
&= 1 - \frac{At}{(tA+1)\binom{d+1+1/t}{d-A+1}}\left[\binom{d+1}{A} + \binom{d+1}{d-A+1}(t(d+1))^{-1} - (At)^{-1}\binom{d}{A-1}\right]\\
&= 1 - \frac{At}{(tA+1)\binom{d+1+1/t}{d-A+1}}\binom{d+1}{A}
\end{aligned}$$

which is the required expression on the RHS. ∎

Acknowledgments. The authors would like to thank the anonymous referees for their remarks and suggestions, which helped in improving the quality of this paper. The first and third authors would also like to thank N.R. Aravind for some initial helpful discussions.

REFERENCES

[1] M. Ajtai, P. Erdős, J. Komlós, and E. Szemerédi, *On Turán's theorem for sparse graphs*, Combinatorica, 1 (1981), pp. 313–317.

[2] M. Ajtai, J. Komlós, and E. Szemerédi, *A note on Ramsey numbers*, J. Combin. Theory Ser. A, 29 (1980), pp. 354–360.

[3] M. Ajtai, J. Komlós, J. Pintz, J.H. Spencer, and E. Szemerédi, *Extremal uncrowded hypergraphs*, J. Combin. Theory Ser. A, 32 (1982), pp. 321–335.

[4] N. Alon and J.H. Spencer, *The Probabilistic Method*, Wiley International, New York, 2008.

[5] Y. Caro, *New Results on the Independence Number*, Technical report, Tel Aviv University, 1979.

[6] Y. Caro and Z. Tuza, *Improved lower bounds on k-independence*, J. Graph Theory, 15 (1991), pp. 99–107.

[7] R. Duke, H. Lefmann, and V. Rödl, *On uncrowded hypergraphs*, Random Structures Algorithms, 6 (1995), pp. 209–212.

[8] R. Bopanna, Comment, *Today is Paul Turan's* 100*th birthday!*, Computational Complexity blog, available from http://blog.computationalcomplexity.org/2010/08/today-is-paul-turans-100th-birthday.html, comment (18 Aug. 2010).

[9] R.L. Graham, D.E. Knuth, and O. Patashnik, *Concrete Mathematics*, Addison-Wesley, Reading, MA, 1994.

[10] J.B. Shearer, *A note on the independence number of triangle-free graphs*, Discrete Math., 46 (1983), pp. 83–87.

[11] J.B. Shearer, *A note on the independence number of triangle-free graphs* II, J. Combin. Theory Ser. B, 2, pp. 300–307.

[12] J.B. Shearer, *On the independence number of sparse graphs*, Random Structures Algorithms, 7 (1995), pp. 269–271.

[13] J. Spencer, *Turán's theorem for k-graphs*, Discrete Math., 2 (1972), pp. 183–186.

[14] R.P. Stanley, *Enumerative Combinatorics*, Cambridge University Press, Cambridge, UK, 1997.

[15] T. Lengyel, *A combinatorial identity and the world series*, SIAM Rev., 35 (1993), pp. 294–297.

Downloaded 02/18/13 to 131.193.153.143. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

[16] T. Thiele, *A lower bound on the independence number of arbitrary hypergraphs*, J. Graph Theory, 30 (1999), pp. 213–221.
[17] P. Turán, *On an extremal problem in graph theory* [in Hungarian], Math Fiz. Lapok, 48 (1941), pp. 436–452.
[18] V.K. Wei, *A Lower Bound on the Stability Number of a Simple Graph*, Technical memorandum TM 81-11217-9, Bell Laboratories, Berkley Heights, NJ, 1981.

Downloaded 02/18/13 to 131.193.153.143. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.