Covering non-uniform hypergraphs

Endre Boros* Yair Caro† Zoltán Füredi‡ Raphael Yuster§

Abstract

A subset of the vertices in a hypergraph is a *cover* if it intersects every edge. Let $\tau(H)$ denote the cardinality of a minimum cover in the hypergraph H, and let us denote by $g(n)$ the maximum of $\tau(H)$ taken over all hypergraphs H with n vertices and with no two hyperedges of the same size. We show that

$$g(n) < 1.98\sqrt{n}(1+o(1)).$$

A special case corresponds to an old problem of Erdős asking for the maximum number of edges in an n-vertex graph with no two cycles of the same length. Denoting this maximum by $n+f(n)$, we can show that $f(n) \leq 1.98\sqrt{n}(1+o(1))$.

Generalizing the above, let $g(n,C,k)$ denote the maximum of $\tau(H)$ taken over all hypergraphs H with n vertices and with at most Ci^k edges with cardinality i for all $i = 1, 2, ..., n$. We prove that

$$g(n,C,k) < (Ck!+1)n^{(k+1)/(k+2)}.$$

These results have an interesting graph-theoretic application. For a family F of graphs, let $T(n,F,r)$ denote the maximum possible number of edges in a graph with n vertices, which contains each member of F at most $r-1$ times. $T(n,F,1) = T(n,F)$ is the classical Turán number. Using the results above, we can compute a non-trivial upper bound for $T(n,F,r)$ for many interesting graph families.

Keywords: Hypergraph, Covering, Cycles
Running Head: Non-Uniform Hypergraphs

1 Introduction

All graphs and hypergraphs considered here are finite, undirected and simple. For the standard terminology used the reader is referred to [3]. Let $H = (V, E)$ be a hypergraph. A subset $T \subset V$ is

*RUTCOR, Rutgers University, 640 Bartholomew Road, Piscataway NJ 08854-8003, U.S.A. e–mail: boros@rutcor.rutgers.edu

†Department of Mathematics, University of Haifa-ORANIM, Tivon 36006, Israel. e–mail: yairc@macam98.ac.il

‡Mathematical Institute of the Hungarian Academy of Sciences, P.O.B. 127, Budapest 1364 Hungary. Research supported in part by the Hungarian National Science Foundation under grant OTKA 016389, and by a National Security Agency grant No. MDA904-98-1-0022. e–mail: furedi@math-inst.hu

§Department of Mathematics, University of Haifa-ORANIM, Tivon 36006, Israel. e–mail: raphy@macam98.ac.il

a *cover* if it intersects all edges, namely $T \cap e \neq \emptyset$ for each $e \in E$. Let $\tau(H)$ denote the cardinality of a minimum cover.

Let $C > 0$ and let k be a nonnegative integer. Let $\mathcal{H}(n, C, k)$ be the family of all hypergraphs with n vertices, having the property that there are at most Ci^k edges with cardinality i, for all $i = 1, 2, ..., n$. In particular, $\mathcal{H}(n) = \mathcal{H}(n, 1, 0)$ denotes the family of all n-vertex hypergraphs whose edges have distinct cardinalities. Put $g(n, C, k) = \max_{H \in \mathcal{H}(n,C,k)} \tau(H)$, and put $g(n) = g(n, 1, 0)$.

Theorem 1.1

$$g(n, C, k) < (Ck! + 1)n^{(k+1)/(k+2)}.$$

Theorem 1.1 gives $g(n) < 2\sqrt{n}$. In this interesting special case we invest some additional effort to improve the upper bound, and supply a lower bound having the same order of magnitude.

Theorem 1.2 *For n sufficiently large $1.5338\sqrt{n} < g(n) < 1.98\sqrt{n}$.*

The determination of $g(n)$ seems to be related to the Turán type problem considered by Chvátal and McDiarmid [7].

The families $\mathcal{H}(n, C, k)$ and $\mathcal{H}(n)$ have interesting graph-theoretic applications. Let F be a family of graphs. Denote by $T(n, F, r)$ the maximum number of edges in a graph on n vertices containing no r isomorphic copies of a member of F. $T(n, F, 1) = T(n, F)$ is just the classical Turán number and is among the most studied parameters in extremal graph theory ([2] Chapter 6 pp. 292-367, [10], [11] Chapter 24 pp.1293-1330). Erdős and Stone [9], and, later, Dirac [8] were the first to raise questions concerning the graphs contained as subgraphs in a graph G on n vertices and $T(n, F) + t$ edges, where t is a positive integer. The Erdős-Stone theorem states, roughly, that with $T(n, K_k)(1 + \epsilon)$ edges one must have not only a copy of K_k but also a copy of the complete k-partite graph with side length $c(k) \log n$, and hence, in particular, many copies of K_k. Dirac's Theorem states that with $T(n, K_k) + 1$ edges there must exist a copy of K_{k+1}^- and hence two copies of K_k. Rademacher ([2] p. 301) posed the specific question of determining the minimum number of triangles in a graph on n vertices and $T(n, K_3) + t$ edges, a problem that was much extended and nearly completely solved years later by Lovász and Simonovits [17].

Our main goal is to present a method to tackle the repeated copies problem in case the growth of $|F_n|$ is bounded from above by a polynomial order, where F_n is the subset of F consisting of graphs with n vertices. We say that F *grows polynomially* if there exist $c > 0$ and a nonnegative integer k such that for every m, there are at most cm^k members in F having exactly m edges. Using Theorem 1.1 and some additional ideas we are able to prove the following theorem:

Theorem 1.3 *Let F be a family of graphs which grows polynomially with parameters c and k. Then, for n sufficiently large,*

$$T(n, F, r) < T(n, F) + (c \cdot (r-1) \cdot k! + 1)T(n, F)^{\frac{k+1}{k+2}} + 2(c \cdot (r-1) \cdot k! + 1)^2 T(n, F)^{\frac{k}{k+2}}. \quad (1)$$

(The constant 2 appearing in front of the final term in (1) can be improved to $1+\epsilon$). There are many interesting families of graphs which grow polynomially. Here are three examples:

- *The family of cycles.* In this case $T(n,F)=n-1$, $c=1$ and $k=0$ as there is only one cycle with m edges for $m\geq 3$. By putting $r=2$ in Theorem 1.3 we get that, for n sufficiently large, every graph with at least $n+2\sqrt{n-1}+7$ edges has two cycles with the same length.

- *The family of subdivisions of a graph.* Let H be any fixed nonempty graph. Recall that a *subdivision* of H is obtained by replacing some (or all) edges of H with paths. Let F_H denote the family of all subdivision of H. For example, F_{K_3} is the family of cycles. If H has h edges, then, clearly, F_H contains at most $\binom{m-1}{h-1}$ graphs with m edges (there may be less, depending on the automorphism group of H). Thus, the family grows polynomially, with $c=1/(h-1)!$ and $k=h-1$. In particular, we have that for n sufficiently large:

$$T(n,F_H,2)<T(n,F_H)+2T(n,F_H)^{\frac{h}{h+1}}+8T(n,F_H)^{\frac{h-1}{h+1}}.$$

  Mader has proved that $T(n,F_H)$ is a linear function of n [18].

- The Family $C(n,t)$ composed of the cycle C_n in which each vertex is also connected to the two vertices at distance t from it on the cycle. Note that the family $C(n,\lfloor n/2\rfloor)$ is rather interesting since for $n\equiv 2 \bmod 4$ it consists of bipartite graphs and hence (1) bounds a non-trivial Turán number.

Erdős (see [3], p. 247) raised the following problem: Let $n+f(n)$ be the maximum number of edges in an n-vertex graph having no two cycles with the same length. Determine $f(n)$. Considering a graph consisting of C_3, $C_4,\ldots$, where C_i is a cycle of length i, and all these cycles have a common vertex but are otherwise pairwise disjoint, one can see that $f(n)\geq\sqrt{2}\sqrt{n}-O(1)$. Using a very similar example Shi [21] proved that $f(n)\geq\lfloor(\sqrt{8n-15}-3)/2\rfloor$ and equality holds for $2\leq n\leq 16$. By giving little more complicated examples, Lai improved the lower bound for $\liminf f(n)/\sqrt{n}$ in a series of notes to $\sqrt{32/15}\sim 1.460\ldots$ [12, 13], then to $\sqrt{162/73}\sim 1.489\ldots$ in [14], and finally to $\sqrt{3249/1381}\sim 1.53383\ldots$ in [15]. Concerning the upper bound, every graph with n vertices contains at least $|E(G)|-n+1$ cycles, hence $f(n)\leq n-3$. However, the order of magnitude of this function is much smaller, as Lai [13] (see also [6]) proved $f(n)\leq O(\sqrt{n\log n})$. As shown above, using Theorem 1.3 we can get $f(n)<2\sqrt{n-1}+7$, thus determining the right order of magnitude of $f(n)$. In fact, we are able to do somewhat better for cycles:

Theorem 1.4 *For n sufficiently large, $f(n)<1.98\sqrt{n}$.*

Combining this with Lai's lower bound we get

$$1.98>\limsup f(n)/\sqrt{n}\geq\liminf f(n)/\sqrt{n}>1.5338.$$

In Section 2 we consider the upper bound for $g(n, C, k)$ and prove Theorem 1.1. The upper and lower bounds for $g(n)$, are handled in Section 3, where we prove Theorem 1.2. Section 3 also considers the fractional covering analog for $g(n)$. Polynomially growing families of graphs, and the proof of Theorem 1.3 appear in Section 4. Section 5 contains the proof of Theorem 1.4. In Section 5 we also consider 2-connected graphs whose cycle lengths are all distinct. We prove that there are such 2-connected graphs with at least $n + \sqrt{n}(1 - o(1))$ edges, improving a result appearing in [6].

2 An upper bound for $g(n, C, k)$

For the proof of Theorem 1.1 let us use some formulations and results from nonlinear binary optimization.

Assume that the set of vertices of the hypergraph H is $[n]$, and let us associate to each subset $S \subset [n]$ its characteristic vector $x^S = (x_1^S, \ldots, x_n^S) \in \{0,1\}^n$ defined by $x_i^S = 1$ iff $i \in S$.

Let us further associate to H a multilinear polynomial $f = f_H$ in n binary variables, defined by

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i + \sum_{e \in H} \prod_{j \in e} (1 - x_j). \tag{2}$$

It is easy to see that $f(x^S) = |S| + t(S)$, where $t(S)$ is the number of edges disjoint from S, and that

$$\tau(H) = \min_{(x_1, x_2, \ldots, x_n) \in \{0,1\}^n} f(x_1, x_2, ..., x_n).$$

Let us also observe that the equality

$$p_i f(p_1, ..., p_{i-1}, 1, p_{i+1}, ..., p_n) + (1 - p_i) f(p_1, ..., p_{i-1}, 0, p_{i+1}, ..., p_n) = f(p_1, ..., p_{i-1}, p_i, p_{i+1}, .., p_n) \tag{3}$$

holds for all $(p_1, ..., p_n) \in [0,1]^n$ and $i = 1, ..., n$, due to the multilinearity of f. Thus

$$\min\{f(p_1, ..., p_{i-1}, 1, p_{i+1}, ..., p_n), f(p_1, ..., p_{i-1}, 0, p_{i+1}, ..., p_n)\} \leq f(p_1, ..., p_{i-1}, p_i, p_{i+1}, .., p_n)$$

follows, implying that a fractional component of a vector can always be switched to an integer value without increasing the value of the function f. By repeating this "rounding" until there are no fractional components, we can arrive from any real vector $(p_1, p_2, ..., p_n) \in [0,1]^n$ to a binary vector $(x_1, x_2, ..., x_n)$ for which we have

$$f(x_1, x_2, ..., x_n) \leq f(p_1, p_2, ..., p_n). \tag{4}$$

In fact, one can implement the above "rounding" procedure to run in $O(n + \sum_{e \in H} |e|)$ time (see e.g. [4, 5]).

As an alternative interpretation, let us consider randomly selected subsets $S \subseteq [n]$, in which the elements are chosen independently with $Prob(i \in S) = p_i$ for $i = 1, ..., n$. Then $Exp[|S| + t(S)] =$

$f(p_1, p_2, ..., p_n)$ follows by simple computation, and hence the existence of a subset S^* for which $f(x^{S^*}) = |S^*| + t(S^*) \leq Exp[|S| + t(S)] = f(p_1, ..., p_n)$ is guaranteed. The above "rounding" procedure therefore can also be viewed as a polynomial (in fact linear) time derandomization of this existential statement (c.f. [1, 20]).

Proof of Theorem 1.1: Let $k \geq 0$ be an integer, let $C > 0$, and let H be a hypergraph with n vertices, where for each i, there are at most Ci^k edges in H having cardinality i. We need to show that

$$\tau(H) < (Ck! + 1)n^{(k+1)/(k+2)}.$$

Let us associate to H the function $f = f_H$ as in (2), and let us consider the subset S, whose characteristic vector we can obtain by the above rounding procedure starting from the real vector $p = (\alpha, \ldots, \alpha)$, for some $0 < \alpha < 1$. According to the above, we have

$$\tau(H) \leq |S| + t(S) = f(x^S) \leq f(\alpha, ..., \alpha),$$

thus it is enough to show that for an appropriate choice of α we have

$$f(\alpha, ..., \alpha) < (Ck! + 1)n^{(k+1)/(k+2)}. \tag{5}$$

Since in the hypergraph H there are at most Ci^k edges having cardinality i for every $i = 1, 2, ..., n$, we get from (2) by simple computation that

$$f(\alpha, ..., \alpha) \leq n\alpha + C\left((1-\alpha) + 2^k(1-\alpha)^2 + \ldots + n^k(1-\alpha)^n\right) < n\alpha + C\sum_{i=1}^{\infty} i^k(1-\alpha)^i. \tag{6}$$

Using the inequality

$$\sum_{i=1}^{\infty} i^k x^i < k!(1-x)^{-k-1}$$

for $0 < x < 1$, which is easy to show by induction on k using term by term derivation, we get from (6) that

$$f(\alpha, ..., \alpha) < n\alpha + Ck!\alpha^{-k-1}.$$

By setting $\alpha = n^{-1/(k+2)}$ yields (5), and hence concludes the proof. $\square$

Let us remark again that the proof of Theorem 1.1 is algorithmic. Namely, given $H \in \mathcal{H}(n, C, k)$ (where C and k are fixed), we can find in polynomial (in n) time a vertex cover whose cardinality is less than the upper bound in the statement of the theorem.

3 Covering hypergraphs whose edge sizes are all distinct

The family $\mathcal{H}(n)$ deserves special attention for two reasons. First, it is a very natural family, consisting of all n-vertex hypergraphs whose edge cardinalities are all distinct. Second, given

Theorem 1.1, we immediately have that $\sqrt{n}$ is the right order of magnitude of $g(n)$, as we have the trivial example of a hypergraph whose edges are the following $(1), (2,3), (4,5,6), (7,8,9,10), \ldots$, giving $g(n) > \sqrt{2n}(1-o(1))$. Thus, it is interesting to close the gap between the upper and lower bounds. Theorem 1.2 improves upon both.

Proof of the upper bound in Theorem 1.2: We need to show that if $H \in \mathcal{H}(n)$, and n is sufficiently large, then $\tau(H) < 1.98\sqrt{n}$. Clearly, we may assume that H has n edges $e_1, \ldots, e_n$ where $|e_i| = i$. We need two lemmas:

Lemma 3.1 *Let F be an n-vertex hypergraph with edges $f_1, \ldots, f_m$ where $|f_i| \le |f_{i+1}|$. Let $a \le m$ be the maximal index for which $|f_1| + \ldots + |f_a| \le n$. Then, $\tau(F) \le (m+a)/2$.*

Proof: Cover $m-a$ edges with at most $(m-a)/2$ vertices, until at most a edges remain uncovered, and then cover each uncovered edge with one vertex. $\square$

Lemma 3.2 *Let $y_1, \ldots, y_n$ be positive reals, and suppose that there is an index γ such that*

$$\sum_{i=1}^{\gamma} i \cdot y_i \le n < \sum_{i=1}^{\gamma+1} i \cdot y_i.$$

Then, conditions $0 \le x_i \le y_i$ for $i = 1, \ldots, n$ and $x_1 + 2x_2 + \ldots + nx_n \le n$ together imply that $x_1 + \ldots + x_n \le y_1 + \ldots + y_{\gamma+1}$.

Proof: Consider the following knapsack problem:

$$\begin{aligned} x_1 + x_2 + \cdots + x_n &\;\to\; \max \\ x_1 + 2x_2 + \cdots + nx_n &\;\le\; n, \text{ and} \\ 0 \le x_i \le y_i &\qquad \text{for } i = 1, 2, ..., n. \end{aligned}$$

It is well known (see e.g. [19]) that for the optimal solution x^* of this knapsack problem we must have

$$x_j^* = \begin{cases} y_j & \text{for } j \le \gamma, \\ 0 & \text{for } j \ge \gamma + 2 \end{cases}$$

and $0 \le x_{\gamma+1}^* \le y_{\gamma+1}$ for some index $0 \le \gamma \le n$. Hence, $x_1^* + \ldots + x_n^* = x_1^* + \ldots + x_{\gamma+1}^* \le y_1 + \ldots + y_{\gamma+1}$ follows, proving the lemma. $\square$

We now proceed with the proof of the upper bound in Theorem 1.2. Assume again that the vertex set of H is $[n]$. Let us consider a random subset X from $[n]$ by including the vertices in X independently with probability p, and let us denote the family of edges disjoint from X by $F(X) = \{f_1, \ldots, f_t\}$, where $t = t(X)$, and where we assume $|f_1| \le |f_2| \le \cdots \le |f_t|$. Let us further denote by $a = a(X)$ (as in Lemma 3.1) the largest index ($\le t(X)$) for which

$$\sum_{j=1}^{a} |f_j| \le n, \tag{7}$$

and let $A(X) = \{f_1, \ldots f_a\}$. Applying Lemma 3.1 to $F(X)$ we obtain

$$\tau(H) \leq |X| + \tau(F(X)) = |X| + \frac{t(X) + a(X)}{2}. \tag{8}$$

To prove the theorem, we shall bound the expected value of the right hand side of (8). First of all we have, as before that

$$\begin{aligned} &Exp[|X|] = np \text{ and} \\ &Exp[t(X)] = \sum_{e \in H} Prob[e \in F(X)] = \sum_{e \in H} (1-p)^{|e|} = \sum_{i=1}^{n} (1-p)^i. \end{aligned} \tag{9}$$

To estimate $Exp[a(X)]$, let us introduce $x_i = Prob(e_i \in A(X))$ for $i = 1, ..., n$. We have

$$x_i \leq (1-p)^i \text{ for } i = 1, ..., n, \tag{10}$$

since $X \cap e_i = \emptyset$ is necessary for $e_i \in A(X)$, and we also have

$$\sum_{i=1}^{n} i x_i \leq n, \tag{11}$$

implied by condition (7) in the definition of $A(X)$. Let us finally define γ as the largest integer $< n$ for which

$$\sum_{i=1}^{\gamma} i(1-p)^i \leq n. \tag{12}$$

Then, by applying Lemma 3.2 with $y_i = (1-p)^i$, we get by (10) and (11) that

$$Exp[a(X)] = \sum_{i=1}^{n} x_i \leq \sum_{i=1}^{\gamma+1} (1-p)^i. \tag{13}$$

Putting together (8) with (9) and (13) we obtain

$$\begin{aligned} \tau(H) &\leq np + \frac{1}{2}\left(\sum_{i=1}^{n} (1-p)^i\right) + \frac{1}{2}\left(\sum_{i=1}^{\gamma+1} (1-p)^i\right) \\ &< np + \frac{1}{2p} + \frac{1}{2}\left(\sum_{i=0}^{\gamma} (1-p)^i\right). \end{aligned} \tag{14}$$

Setting $p = \frac{\alpha}{\sqrt{n}}$ for some $0 < \alpha < 1$, and introducing $1 < \beta \leq \sqrt{n}$ such that $\gamma + 1 = \beta\sqrt{n}$, and using that for large enough n we have $(1 - \frac{\alpha}{\sqrt{n}})^{\beta\sqrt{n}} \approx e^{-\alpha\beta}$, we get from (14) that

$$\tau(H) < \alpha\sqrt{n} + \frac{\sqrt{n}}{2\alpha} + \frac{1}{2}\frac{\sqrt{n}}{\alpha}(1 - \exp(-\alpha\beta) + o(1)).$$

Hence,

$$\frac{\tau(H)}{\sqrt{n}} < \alpha + \frac{1}{\alpha} - \frac{\exp(-\alpha\beta)}{2\alpha} + o(1). \tag{15}$$

On the other hand, the definition of β and (12) implies that

$$\frac{1}{\alpha^2} - \frac{1+\alpha\beta}{\alpha^2}\exp(-\alpha\beta) = 1 + o(1). \tag{16}$$

We used here the fact that

$$\sum_{i=1}^{\gamma} ix^i = x\frac{(x-1)(\gamma+1)x^\gamma + 1 - x^{\gamma+1}}{(1-x)^2}.$$

Minimizing the r.h.s. of (15) subject to (16) we get that $\alpha = 0.808\ldots$ and $\beta = 2.760\ldots$ yielding $1.97913\ldots$ for the r.h.s. of (15). Hence, for n sufficiently large, $\tau(H) < 1.98\sqrt{n}$. $\square$

Proof of the lower bound in Theorem 1.2: The proof of the lower bound is facilitated by the following simple lemma:

Lemma 3.3 *Let F be a polynomially growing family of graphs with parameters c and k. Then,*

$$T(n,F,r) \le T(n,F) + g(T(n,F,r), c\cdot(r-1), k).$$

In particular, $f(n) \le g(n+f(n)) - 1$.

Proof: Consider a graph G with n vertices and with $T(n,F,r)$ edges, in which each member of F appears as a subgraph of G at most $r-1$ times. We create a hypergraph H whose vertices are the edges of G and whose edges correspond to the edge sets of subgraphs of G which are isomorphic to some member of F. Clearly, $H \in \mathcal{H}(T(n,F,r), c\cdot(r-1), k)$. Thus, there is a subset of at most $g(T(n,F,r), c\cdot(r-1), k)$ edges of G whose deletion from G makes it F-free. Thus, $T(n,F,r) - g(T(n,F,r), c\cdot(r-1), k) \le T(n,F)$. The fact that $f(n) \le g(n+f(n)) - 1$ follows by observing that if F is the family of cycles then $n + f(n) = T(n,F,2)$, $c = 1$, $k = 0$, and $T(n,F) = n-1$. $\square$

By Lemma 3.3 we get $g(n+f(n)) \ge f(n)+1$. As noted in the introduction, Lai has shown, that for every sufficiently large n, $f(n) > 1.53383\sqrt{n}$. Hence, $g(n+1.53383\sqrt{n}) \ge 1.53383\sqrt{n}+1$. This, implies that $g(n) > 1.5338\sqrt{n}$ for n sufficiently large. $\square$

In fact, Lemma 3.3 shows that $\liminf g(n)/\sqrt{n} \ge \liminf f(n)/\sqrt{n}$. It may be that the inequality is strict, since the hypergraphs in $\mathcal{H}(n)$ do not need to possess any structure, and, in particular, they may not have a back translation to a graph whose cycle lengths are all distinct. The authors have, in fact, a construction which shows $\liminf g(n)/\sqrt{n} \ge \sqrt{22/9} = 1.5634$. We omit the details.

An assignment of nonnegative weights to the vertices of a hypergraph is a *fractional cover* if the sum of the weights of the vertices of each edge is at least 1. Let $\tau^*(H)$ be the smallest possible sum of weights of a fractional cover. Clearly, $\tau^*(H) \le \tau(H)$. Let $g^*(n)$ denote the maximum value of $\tau^*(H)$ taken over all graphs in $\mathcal{H}(n)$. Clearly, $g^*(n) \leq g(n)$. Ron Holzman suggested that the determination of $g^*(n)$ might be easier than that of $g(n)$. This is indeed the case, as shown in the following proposition:

Proposition 3.4 $g^*(n) = \sqrt{2n} + O(1)$.

Proof: Suppose $r(r+1)/2 \leq n < (r+1)(r+2)/2$. The lower bound $g^*(n) \geq r$ is obtained from r disjoint sets. For the upper bound, assign a weight of $1/(r+1)$ to each vertex. Every edge with at least $r+1$ vertices is covered. Add $(r+1-|e|)/(r+1)$ additional weight to each small edge e. This gives $\tau(H) \leq n/(r+1) + r/2$. In fact, we have shown $g^*((r^2+r)/2) = r$. $\square$

The greedy algorithm for a cover in a hypergraph is defined as follows. At each stage, pick a vertex of maximum degree and delete all edges incident with that vertex. Continue in the same manner until all edges are covered. Let $g'(n)$ denote the maximum size of a cover produced by the greedy algorithm, where the maximum is taken over all graphs in $\mathcal{H}(n)$. The following proposition shows that the greedy algorithm produces a relatively small covering:

Proposition 3.5 $g'(n) \leq 2.7\sqrt{n}$.

Proof: Let $g'(n,r)$ be the maximum size of a vertex cover produced by the greedy algorithm, where the maximum is taken over all hypergraphs in $\mathcal{H}(n)$ having precisely r edges. Clearly, $g'(n,r) = r$ whenever $r(r+1)/2 \leq n$. Considering the average degree we get that

$$g'(n,r) \leq 1 + g'(n-1, r - \lceil \frac{r(r+1)}{2n} \rceil).$$

One can now prove by induction that

$$g'(n) = g'(n,n) \leq \sqrt{2n} \sum_{i=0}^{n} \frac{\sqrt{i+1} - \sqrt{i}}{i+1} < 2.7\sqrt{n}. \qquad \square$$

It is interesting to find a nontrivial lower bound for $g'(n)$, namely, one that is significantly larger than the lower bound for $g(n)$.

4 An upper bound for $T(n,F,r)$

Proof of Theorem 1.3: Let G have n vertices and m edges where

$$m = \lceil T(n,F) + (c \cdot (r-1) \cdot k! + 1) T(n,F)^{\frac{k+1}{k+2}} + 2(c \cdot (r-1) \cdot k! + 1)^2 T(n,F)^{\frac{k}{k+2}} \rceil.$$

We need to show that there exists some member of F which appears at least r times in G. As in the proof of Lemma 3.3, it suffices to show that $m - g(m, c \cdot (r-1), k) > T(n,F)$. By Theorem 1.1 it suffices to show that

$$m > T(n,F) + (c \cdot (r-1) \cdot k! + 1) m^{(k+1)/(k+2)}.$$

By the definition of m, it suffices to show that for n sufficiently large:

$$(c \cdot (r-1) \cdot k! + 1)T(n,F)^{\frac{k+1}{k+2}} + 2(c \cdot (r-1) \cdot k! + 1)^2 T(n,F)^{\frac{k}{k+2}} >$$

$$(c \cdot (r-1) \cdot k! + 1)(T(n,F) + (c \cdot (r-1) \cdot k! + 1)T(n,F)^{\frac{k+1}{k+2}} + 2(c \cdot (r-1) \cdot k! + 1)^2 T(n,F)^{\frac{k}{k+2}})^{\frac{k+1}{k+2}}.$$

To simplify notation put $D = c(r-1)k! + 1$, $T = T(n,F)$ and $\beta = (k+1)/(k+2)$. We must show:

$$DT^\beta + 2D^2T^{2\beta-1} > D(T + DT^\beta + 2D^2T^{2\beta-1})^\beta.$$

Dividing by DT^β the last inequality is equivalent to:

$$1 + 2DT^{\beta-1} > (1 + DT^{\beta-1} + 2D^2T^{2\beta-2})^\beta.$$

Since $\beta < 1$ it suffices to show that

$$1 + 2DT^{\beta-1} > 1 + DT^{\beta-1} + 2D^2T^{2\beta-2}.$$

The last inequality is equivalent to

$$T^{1-\beta} > 2D$$

which clearly holds for n (and, thus, also $T = T(n,F)$) sufficiently large. $\square$

5 Graphs whose cycles have distinct lengths

Proof of Theorem 1.4: According to Lemma 3.3 $f(n) \le g(n + f(n)) - 1$. By Theorem 1.2, $g(n) < 1.98\sqrt{n}$, (in fact 1.97914, as shown in the proof) if n is sufficiently large. Thus, $f(n) < 1.97914\sqrt{n + f(n)} - 1$. Hence, $f(n) < 1.98\sqrt{n}$ for n sufficiently large. $\square$

The proof of Theorem 1.2, however, does not assume any structure of the hypergraph in question. However, if the hypergraph H is formed, as in Lemma 3.3, from a graph G whose cycle lengths are all distinct, then there is a structure imposed. This structure enables us to slightly improve upon the 1.98 upper bound. More precisely, after selecting the random set X in the proof of Theorem 1.1 we proceed as follows. If we find a vertex of H incident with at least three remaining edges of H, we pick it for the cover and by that we eliminate at least three edges. We continue doing so until every vertex is on one or two edges. This means that in G, the 2-connected components of the remaining edges are either cycles or Θ-graphs. Consequently, this means that the edges of H can be partitioned into blocks where each block contains either a single edge (whose vertices appear nowhere else; this corresponds to a 2-connected component of G which is a cycle), or three edges where every vertex which appears in one of the three, appears also in another one, and nowhere else (this corresponds to a 2-connected component of G which is a Θ-graph). Utilizing this special structure we can get a bound of $1.945\sqrt{n}$. We omit the precise details.

For completeness, here is a simple construction which shows $\liminf f(n)/\sqrt{n} \geq \sqrt{7/3} = 1.527\ldots$. Although slightly less than Lai's lower bound of $1.5338\ldots$, this construction is very simple. We show that for any integer $k \geq 1$ there exists a graph G^k on $n = 21k^2 - 4k$ vertices with $n + 7k - 2$ edges containing all cycles of lengths from 3 to $10k$ exactly once. G^k consists of 2-connected blocks, B_ℓ, $3 \leq \ell \leq 5k$. These blocks all have a common vertex x, otherwise their vertex sets are pairwise disjoint. For $\ell \leq 4k$ the block B_ℓ is simply a cycle of length ℓ. For $\ell = 4k + 1 + i$, $0 \leq i \leq k - 1$, the block B_ℓ is obtained by taking two cycles $C_{4k+2i+1}$ and $C_{4k+2i+2}$ with a single common vertex x, taking points $x_{4k+2i+1} \in V(C_{4k+2i+1})$, $x_{4k+2i+2} \in V(C_{4k+2i+2})$ such that their distance from x is exactly $2k + i$ and, finally, connecting $x_{4k+2i+1}$ and $x_{4k+2i+2}$ by a new path consisting of $2k + 2i + 1$ edges. This block has $10k + 6i + 4$ edges and contains six cycles of lengths $4k + 2i + 1$, $4k + 2i + 2$ and $6k + 4i + \alpha$, $\alpha = 1, 2, 3, 4$.

Let $f_2(n) + n$ be the maximum number of edges in a simple, 2-connected graph on n vertices with the property that any two cycles have distinct lengths. Shi [21] proved, using the well-known ear-decomposition, that every 2-connected graph with n vertices and $n + b$ edges contains at least $\binom{b+2}{2}$ cycles. This implies $f_2(n) \leq (\sqrt{8n - 15} - 3)/2 \sim \sqrt{2}\sqrt{n}$. On the other hand, it is shown in [6], that $f_2(n) \geq \sqrt{n/2}(1 - o(1))$. In the following Proposition we improve this lower bound significantly:

Proposition 5.1 $f_2(n) \geq \sqrt{n} - O(n^{9/20})$.

Proof: A sequence of integers, $a_1, \ldots, a_k$, forms a *Sidon* sequence if all the $\binom{k+1}{2}$ sums of the form $a_i + a_j$ (where $1 \leq i \leq j \leq k$) are distinct. Let $b_2(n)$ denote the size of the largest Sidon subsequence of $[n]$. An old Theorem of Erdős and Turán [16] states, that $b_2(n) \sim \sqrt{n}$. The lower bound in this theorem is supplied by Singer's Theorem [22], which states: For every prime power p there exists a sequence of integers $a_1, a_2, \ldots, a_{p+1}$, such that the $(p+1)p$ differences $a_i - a_j$ $(i \neq j)$ produce all the numbers $1, 2, \ldots, p(p+1)$ modulo $p^2 + p + 1$. Such a sequence is called a *difference set* mod $p^2 + p + 1$.

Now let p be a prime $\sqrt{n} < p < \sqrt{n} + O(n^{2/5})$, so for $m = p^2 + p + 1$ we have $m - n = O(n^{9/10})$. Let $a_1, a_2, \ldots, a_{p+1}$ be a difference set modulo m. There exists a unique solution of the equation $a_i + (n-2) \equiv a_j \pmod{m}$. Observe, that for any integer r the sequence $\{a_i + r\}$ is a difference set, too (addition is mod m). So we may suppose, that after an appropriate shifting, $a_1 = 1 < a_2 < \cdots < a_k = n - 1 < \cdots < a_{p+1} \leq m$. We have that $p + 1 - k \leq b_2(m - (n-1)) = (1 + o(1))\sqrt{m - n} = O(n^{9/20})$, by the Erdős Turán Theorem.

We construct a 2-connected graph G as follows. $E(G)$ consists of a Hamilton cycle $\{v_0, v_1, \ldots, v_{n-1}\}$ (in this order) and the edges $v_0 v_{a_i}$, $1 < i < k$. By construction, it has no two cycles with the same length and has $n + k - 2 = n + \sqrt{n} - O(n^{9/20})$ edges. □

Combining the last proposition with Shi's result we get:

Corollary 5.2 $\sqrt{2} \geq \limsup f_2(n)/\sqrt{n} \geq \liminf f_2(n)/\sqrt{n} \geq 1$.

We make the following Conjecture:

Conjecture 5.3 $\lim f_2(n)/\sqrt{n} = 1$.

It is easy to see that Conjecture 5.3 implies the (difficult) upper bound in the Erdős Turán Theorem.

Acknowledgment

The authors are indebted to Janice Malouf for fruitful discussions. The first author also gratefully acknowledges the partial support by the Office of Naval Research (Grant N00014-92-J-1375) and by the National Science Foundation (Grant DMS 98-06389).

References

[1] Alon, N., and Spencer, J.H., *The probabilistic method*, John Wiley and Sons, New York, 1992.

[2] B. Bollobás, *Extremal Graph Theory*, Academic Press, 1978.

[3] J. A. Bondy and U.S. R. Murty, *Graph Theory with Applications*, Macmillan Press, London, 1976.

[4] E. Boros and A. Prékopa, *Probabilistic bounds and algorithms for the maximum satisfiability problem*, Annals of Operations Research 21 (1989), pp. 109–126.

[5] E. Boros, *Maximum renamable Horn sub-CNFs*, Discrete Applied Mathematics 96-97 (1999), pp. 29–40.

[6] G. Chen, J. Lehel, M.S. Jacobson and W.E. Shreve, *Note on graphs without repeated cycle length*, J. Graph Theory 29 (1998), 11-15.

[7] V. Chvátal and C. McDiarmid, *Small transversals in hypergraphs*, Combinatorica 12 (1992), 19–26.

[8] G.A. Dirac, *Extensions of Turán's theorem on graphs*, Acta Math. Sci. Hungar. 14 (1963), 418-422.

[9] P. Erdős and A.H. Stone *On the structure of linear graphs*, Bull. Amer. Math. Soc. 52 (1946), 1087-1091.

[10] Z. Furedi, *Turán Type Problems*, In: Surveys in Combinatorics, London Mathematical Society, lecture note series 166 (A.D. Keedwell ed.), 1991.

[11] R.L. Graham, M. Grotschel and L. Lovász, *Handbook of Combinatorics*, North-Holland publishers, 1995.

[12] Chun Hui Lai, *On a problem of Erdős*, (Chinese), Zhangzhou Shiyuan Xuebao (Ziran Kexue Ban) (1989), No. 1, 55-59. MR91a:05063.

[13] Chun Hui Lai, *Upper and lower bounds for* $f(n)$, (Chinese), Zhangzhou Shiyuan Xuebao (Ziran Kexue Ban) 4 (1990), No. 1, 30-34, 29.

[14] Chun Hui Lai, *On the size of graphs with all cycle having distinct length*, Discrete Mathematics 122 (1993), 363-364.

[15] Chun Hui Lai, *The number of edges of graphs in which all cycles have different lengths*, (Chinese), Zhangzhou Shiyuan Xuebao (Ziran Kexue Ban) 8 (1994), No. 4, 30–34. MR96h:05121.

[16] P. Erdős and P. Turán, *On a problem of Sidon in additive number theory, and some related problems*, J. London MathṠoc1̇6 (1941), 212–215. Addendum (by P. Erdős) ibid1̇9 (1944), 208.

[17] L. Lovász and M.Simonovits, *On the number of complete subgraphs of a graph*, Proc. 5^{th} British Combinatorial Conference (C. Nash-Williams, J. Sheehan eds.), Utilitas Math. Winnipeg (1976), 431-441.

[18] W. Mader, *Hinreichende bedingungen für die existenze von teilgraphen, die zu einem vollständigen graphen homöomorph sind*, Math. Nachr. 53 (1972), 145-150.

[19] G. L. Nemhauser and L. A. Wolsey, *Integer and Combinatorial Optimization*, J. Wiley, 1988.

[20] Raghavan, P., and Thompson, C., *Randomized rounding: A technique for provably good algorithms and algorithmic proofs*, Combinatorica 7 (1987), pp. 365-374.

[21] Yongbing Shi, *On maximum cycle-distributed graphs*, Discrete Math. 71 (1988), 57-71.

[22] J. Singer, *A theorem in finite projective geometry and some applications to number theory*, Trans. Amer. Math. Soc. 43 (1938), 377-385.