Improved Bounds For Covering Complete Uniform Hypergraphs

Jaikumar Radhakrishnan, Department of Computer Science,

Rutgers University, New Brunswick, NJ 08903, USA

April 3, 2007

Abstract

We consider the problem of covering the complete r-uniform hypergraphs on n vertices using complete r-partite graphs. We obtain lower bounds on the size of such a covering. For small values of r our result implies a lower bound of $\Omega(\frac{e^r}{r\sqrt{r}} n \log n)$ on the size of any such covering. This improves the previous bound of $\Omega(rn \log n)$ due to Snir [5]. We also obtain good lower bounds on the size of a family of perfect hash function using simple arguments.

Key words. Combinatorial problems, graph covering, perfect hashing, graph entropy.

1 Introduction

Let r and n be positive integers such that $r \leq n$. Let $[n] = \{1, 2, \ldots, n\}$. Let V be a finite set. Let $\binom{V}{r} = \{T \subseteq V : |T| = r\}$. Let $(n)_r = n(n-1)\ldots(n-r+1)$.

A *hypergraph* H is a pair $(V(H), E(H))$ where $V(H)$ is the set of vertices of H and $E(H)$ is a collection of subsets of $V(H)$. A hypergraph H is said be *r-uniform* if all the elements of $E(H)$ have size r. Thus an r-uniform hypergraph H with vertex set $[n]$ satisfies $E(H) \subseteq \binom{[n]}{r}$. If

$E(H) = \binom{V(H)}{r}$ then H is said to be a complete r-uniform hypergraph. We shall use $K_n(r)$ to denote the complete r-uniform hypergraph with vertex set $[n]$. We shall refer to 2-uniform hypergraphs simply as *graphs*.

A *subpartition* of $[n]$ is a set of pairwise disjoint subsets of $[n]$. The degree of a subpartition is the number of subsets in it. The size of a subpartition $\mathcal{A}$, denoted by $S(\mathcal{A})$, is the sum of the sizes of the subsets appearing in $\mathcal{A}$. Let $\mathcal{A} = \{A_1, A_2, \ldots, A_r\}$ be a subpartition of $[n]$. Then $K_n(\mathcal{A})$, the hypergraph induced by $\mathcal{A}$, is the r-uniform hypergraph such that $V(K_n(\mathcal{A})) = [n]$ and $E(K_n(\mathcal{A})) = \{T \in \binom{[n]}{r} : |T \cap A_j| = 1 \textit{ for } j = 1, \ldots, r\}$.

For hypergraphs H_1 and H_2 with $V(H_1) = V(H_2)$, $H_1 \cup H_2$ will be the hypergraph such that $V(H_1 \cup H_2) = V(H_1)$ and $E(H_1 \cup H_2) = E(H_1) \cup E(H_2)$.

For $i = 1, \ldots, h$, let $\mathcal{A}^i$ be a subpartition of $[n]$ of degree r. We say that the family $\Gamma = \{K_n(\mathcal{A}^1), K_n(\mathcal{A}^2), \ldots, K_n(\mathcal{A}^h)\}$ is a *covering* of $K_n(r)$ if $K_n(r) = \bigcup_{i=1}^h K_n(\mathcal{A}^i)$. Let $S(\Gamma) = \sum_{i=1}^h S(\mathcal{A}^i)$. Let $g_r(n)$ be the minimum $S(\Gamma)$ over all coverings Γ of $K_n(r)$.

In this note we will be concerned with showing good lower bounds for $g_r(n)$. A straight forward counting argument gives that

$$g_r(n) \geq n \binom{n}{r} \left(\frac{n}{r}\right)^{-r}. \tag{1}$$

By choosing random subpartitions (see [1], [2]) one can show

$$g_r(n) = O\left(\frac{(n)_r}{n^r} \frac{r^r}{r!} n \log \binom{n}{r}\right). \tag{2}$$

If r does not grow faster then $\log n$, (2) implies

$$g_r(n) = O(\sqrt{r} \exp(r) n \log n).$$

On the other hand, for these values of r, the lower bound given by (1),

$$n\binom{n}{r}\left(\frac{n}{r}\right)^{-r} = O(\frac{\exp(r)}{\sqrt{r}}\, n).$$

In this note, we apply Körner's technique [2] to this problem and show, for $2 \leq r \leq n$,

$$g_r(n) \geq \frac{(n)_{r-1}}{n^{r-1}} \frac{r^{r-1}}{(r)_{r-1}}\, n \log(n - r + 2).$$

For the values of r less than $\log n$, this gives

$$g_r(n) = \Omega(\frac{\exp(r)}{r\sqrt{r}} n \log n). \tag{3}$$

The previous bound due to Snir [5] gives, for $2 \leq r \leq n$,

$$g_r(n) \geq n\, \frac{\log n - \log(r-1)}{\log r - \log(r-1)}.$$

Note that $n\, \frac{\log n - \log(r-1)}{\log r - \log(r-1)}$=O$(rn \log n)$.

This problem is related to computation of threshold functions using certain restricted kinds of formulas called $\Sigma\Pi\Sigma$ formulas. $\Sigma\Pi\Sigma$ formulas have the form $\bigvee_{i=1}^{p} \bigwedge_{j=1}^{t_i} \bigvee_{q \in S_{ij}} q$, where each S_{ij} is a subset of variables and their negations. The size of such a formula is the sum of the sizes of the S_{ij}. The threshold function T_k^n is the boolean function on n variables that takes the value 1 precisely when there are at least k 1's in the input.

Newman, Ragde, and Wigderson [3] observed that the problem of determining the size of the smallest $\Sigma\Pi\Sigma$ formula computing T_k^n is equivalent to the hypergraph covering problem stated above, if the t_i are restricted to be k (the fanin of the AND gates is restricted to be k). Using the the notion of hypergraph entropy they restated Snir's result. However, if the fanin of the AND gates is not restricted to be k, the two problems are not equivalent. Indeed, there exist $\Sigma\Pi\Sigma$ formulas for computing T_k^n, for $k \leq \log n$, with size $O(\exp(3\sqrt{k}\log k)n\log n)$ (see [4]). This is better than the lower bound given by (3) above.

A related problem arises in the study of families of perfect hash functions. Let X be an n element set. Let B be a b element set. Following Körner [2] we say that a function $f : X \to B$ *separates* the set $A \subseteq X$ if f takes a different value on each element of A. Let f_π, $\pi \in \Pi$ be a family of mappings of the set X into a set B. The family $\{f_\pi : \pi \in \Pi\}$ is said to be a (b,k)-*family* of perfect hash functions for X if every k-element subset of X is separated by at least one function f_π, $\pi \in \Pi$. Let $Y(b,k,n)$ be the minimum size of any (b,k)-family of perfect hash functions for the set $[n]$.

Fredman and Komlós [1] obtained good lower bounds for $Y(b,k,n)$ using an information inequality concerning a measure on graphs. Later Körner restated the Fredman-Komlós bound using a graph entropy. Their results show, for b and k fixed with respect to n, asymptotically

$$\frac{b^{k-1}}{(b)_{k-1}}\frac{\log n}{\log(b-k+2)} \leq Y(b,k,n) \leq \frac{b^k}{(b)_k}\,k\log n.$$

In section 4, we shall show lower bounds for $Y(b,k,n)$ that come close to the Fredman-Komlós bound. However, our simpler argument will use only the elementary fact that a complete graph on n vertices can not be expressed as a union of less than $\frac{\log n}{\log r}$ r-partite graphs.

2 Graph Entropy

In this section we review the basic facts about entropy of graphs. The following definitions and results are from Körner [2]. All logarithms in this paper are with base 2.

Definition 2.1 (Entropy) Given a random variable X with finite range, its entropy is given by

$$\mathrm{H}(X) = -\sum_{x} \Pr[X = x] \log \Pr[X = x].$$

Definition 2.2 (Mutual Information) If X and Y are random variables with finite ranges, then their mutual information is given by $\mathrm{I}(X \wedge Y) = \mathrm{H}(X) + \mathrm{H}(Y) - \mathrm{H}((X, Y))$.

Definition 2.3 (Graph Entropy) Let $G = (V, E)$ be a graph. Let P be a probability distribution on the vertex set V. Let $\mathcal{A}(G)$ denote the set of all independent sets of G. Let $\mathcal{P}(G)$, the set of admissible distributions, be the set of all distributions Q_{XY} on $V \times \mathcal{A}(G)$ satisfying (a) $\mathrm{Q}_{XY}(v, A) = 0$ if $v \notin A$, and (b) $\sum_A \mathrm{Q}_{XY}(v, A) = \mathrm{P}(v)$ for all vertices v in V. The graph entropy $\mathrm{H}(G, P)$ is defined by

$$\mathrm{H}(G, P) = \min\{\mathrm{I}(X \wedge Y) : \mathrm{Q}_{XY} \in \mathcal{P}(G)\}.$$

Lemma 2.4 (Subadditivity of Graph Entropy) If G and F are graphs with $V(G) = V(F)$, and P is a distribution on $V(G)$, then $\mathrm{H}(F \cup G, P) \leq \mathrm{H}(F, P) + \mathrm{H}(G, P)$.□

In our discussion P will always be assumed to be the uniform distribution and will be omitted from our notation for graph entropy. It is easy to see that under this condition the entropy of the complete graph on n vertices is $\log n$. The entropy of the empty graph is 0. The entropy of a complete bipartite graph is at most 1.

Lemma 2.5 (Additivity of Graph Entropy) Let $\{G_i\}_{i\in I}$ be the set of connected components of a graph G. Then

$$\mathrm{H}(G) = \sum_{i\in I} \frac{|V(G_i)|}{|V(G)|}\mathrm{H}(G_i). \ \square$$

3 The Lower Bound

The idea of the lower bound is as follows. We associate with each r-uniform hypergraph a simple graph (its Fredman-Komlós graph). The simple graph associated with $K_n(r)$ has high entropy. Under this association the graph associated with $K_n(r)$ will be the union of the graphs associated with the r-partite hypergraphs in its covering. The graphs associated with the r-partite hypergraphs will have low entropy. Our lower bound result will then follow by the subadditivity of graph entropy. We now present our argument in detail.

Definition 3.1 (Fredman-Komlós Graph) Let $r \geq 2$. Let H be an r-uniform hypergraph on vertex set $[n]$. Then $G(H, r)$ is the graph defined by

$$\begin{aligned} V(G(H,r)) &= \{(C,x) : C \in \binom{[n]}{r-2} \text{ and } x \in [n] - C\}; \\ E(G(H,r)) &= \{((C,x),(D,y)) : C = D \text{ and } C \cup \{x,y\} \in E(H)\}. \end{aligned}$$

Thus $G(K_n(r), r)$ consists of $\binom{n}{r-2}$ components, where each component is a complete graph on $n - r + 2$ vertices. In general, if H is any r-uniform hypergraph then the subgraph of $G(H, r)$ induced by those vertices (C, x) that have the same value for C will be called a *block* of $G(H, r)$. Thus each block has $n - r + 2$ vertices and there are $\binom{n}{r-2}$ blocks, one for each $C \in \binom{[n]}{r-2}$. Every edge of $G(H, r)$ is contained in one of its blocks. The following lemma is a direct consequence of lemma 2.5 and the earlier observation about the entropy of complete graphs.

Lemma 3.2 $H(G(K_n(r), r)) = \log(n - r + 2)$. $\square$

Let $\mathcal{A}$ be a subpartition of $[n]$ of degree r. Consider $G(K_n(\mathcal{A}), r)$. Suppose that the block corresponding to $C = \{c_1, c_2, \ldots, c_{r-2}\}$ is non-empty. Then there exists a set $T = \{A_1, A_2, \ldots, A_{r-2}\} \in \binom{\mathcal{A}}{r-2}$ such that $c_1 \in A_1, c_2 \in A_2, \ldots, c_{r-2} \in A_{r-2}$. Further, the edges within this block are arranged as a bipartite graph. Indeed, if $\mathcal{A} - T = \{A_{r-1}, A_r\}$, then the edges within this block are of the form $((C, x), (C, y))$ where $x \in A_{r-1}$ and $y \in A_r$. Since the entropy of a bipartite graph is at most 1 and since all but $|A_{r-1} \cup A_r|$ of the vertices are isolated, we get (using lemma 2.5) that the entropy of the block is at most $\frac{1}{n-r+2}|A_{r-1} \cup A_r|$.

Lemma 3.3 *If $\mathcal{A}$ is a subpartition of $[n]$ of degree r, then*

$$H(G(K_n(\mathcal{A}), r)) \leq \frac{S(\mathcal{A})}{n} \frac{n^{r-1}}{(n)_{r-1}} \frac{(r)_{r-1}}{r^{r-1}}.$$

Proof : As described above, to each non-empty block of $G(K_n(\mathcal{A}), r)$ there corresponds a set of $T \in \binom{\mathcal{A}}{r-2}$. The number of non-empty blocks that correspond to a set T is $\prod_{A \in T} |A|$. The entropy of each of these blocks is at most $\frac{1}{n-r+2}|A_1 \cup A_2|$, where $\{A_1, A_2\} = \mathcal{A} - T$. We may thus conclude using lemma 2.5 that

$$H(G(K_n(\mathcal{A}), r)) \leq \frac{1}{\binom{n}{r-2}} \sum_{T \in \binom{\mathcal{A}}{r-2}} (\prod_{A \in T} |A|) \frac{1}{n - r + 2} (\sum_{B \in \mathcal{A} - T} |B|).$$

For $S(\mathcal{A})$ fixed, the expression on the right is maximized if $|A| = \frac{S(\mathcal{A})}{r}$ for each $A \in \mathcal{A}$. Thus

$$\begin{aligned} H(G(K_n(\mathcal{A}), r)) &\leq \frac{2}{(n-r+2)\binom{n}{r-2}} \binom{r}{r-2} \left(\frac{S(\mathcal{A})}{r}\right)^{r-1} \\ &= \left(\frac{S(\mathcal{A})}{n}\right)^{r-1} \frac{n^{r-1}}{(n)_{r-1}} \frac{(r)_{r-1}}{r^{r-1}} \\ &\leq \frac{S(\mathcal{A})}{n} \frac{n^{r-1}}{(n)_{r-1}} \frac{(r)_{r-1}}{r^{r-1}}. \end{aligned}$$

The last inequality holds because $S(\mathcal{A}) \leq n$ and $r \geq 2$. □

We are now ready to prove our main result.

Theorem 3.4 *If* $2 \leq r \leq n$*, then*

$$g_r(n) \geq \frac{(n)_{r-1}}{n^{r-1}} \frac{r^{r-1}}{(r)_{r-1}} n \log(n-r+2).$$

Proof : Let $\mathcal{A}^1, \ldots, \mathcal{A}^h$ be subpartitions of $[n]$ of degree r. Let $\Gamma = \{K_n(\mathcal{A}^1), K_n(\mathcal{A}^2), \ldots, K_n(\mathcal{A}^h)\}$ be a covering for $K_n(r)$. Then

$$G(K_n(\mathcal{A}^1), r) \cup G(K_n(\mathcal{A}^2), r) \cup \ldots \cup G(K_n(\mathcal{A}^h), r) = G(K_n(r), r).$$

By lemma 2.4 we have that

$$H(G(K_n(\mathcal{A}^1), r)) + H(G(K_n(\mathcal{A}^2), r)) + \ldots + H(G(K_n(\mathcal{A}^h), r)) \geq H(G(K_n(r), r)).$$

By lemma 3.2 and lemma 3.3 we have

$$\sum_{i=1}^{h} \frac{S(\mathcal{A}^i)}{n} \frac{n^{r-1}}{(n)_{r-1}} \frac{(r)_{r-1}}{r^{r-1}} \geq \log(n-r+2).$$

It follows that

$$S(\Gamma) = \sum_{i=1}^{h} S(\mathcal{A}^i) \geq \frac{(n)_{r-1}}{n^{r-1}} \frac{r^{r-1}}{(r)_{r-1}} n \log(n-r+2).$$

The proof of the theorem is complete. □

4 The Fredman-Komlós Bound

In this section we describe the lower bound for $Y(b,k,n)$.

Theorem 4.1 *For b and k fixed and n large enough*

$$Y(b,k,n) \geq \frac{b^{k-2}}{(b)_{k-2}} \frac{\log(n-k+2)}{\log(b-k+2)}.$$

We shall make use of the following elementary fact: *A complete graph on n vertices can not be expressed as a union of less than $\frac{\log n}{\log r}$ r-partite graphs.*

The following definition associates each hash function with a graph.

Definition 4.2 (Fredman-Komlós Graph) *Let $f : [n] \to B$ be a hash function. The k-th Fredman-Komlós graph for f, denoted by $G(f,k)$, is defined by*

$$V(G(f,k)) = \{(C,i) : C \in \binom{[n]}{k-2} \text{ and } i \in [n] - C\};$$

$$E(G(f,k)) = \{((C,i),(D,j)) : C = D, |C \cup \{i,j\}| = k \text{ and } f \text{ separates } C \cup \{i,j\}\}.$$

The k-th complete Fredman-Komlós graph, denoted by C(k), is defined by

$$V(C(k)) = \{(C,i) : C \in \binom{[n]}{k-2} \text{ and } i \in [n] - C\};$$
$$E(G(f,k)) = \{((C,i),(D,j)) : C = D, |C \cup \{i,j\}| = k\}.$$

Proof of the theorem : Let $\{f_\pi : \pi \in \Pi\}$ be a minimum size (b,k)-family of perfect hash functions for n. It is easy to see that

$$\bigcup_{\pi \in \Pi} G(f_\pi, k) = C(k).$$

Now, $C(k)$ consists of $\binom{n}{k-2}$ components each of which is a complete graph on $n-k+2$ vertices. On the other hand, the contribution of a $G(f_\pi,k)$ to each of these components is either a complete $(b-k+2)$-partite graph or nothing at all. The number of components of $C(k)$ to which the contribution of a $G(f_\pi,k)$ is non-empty is at most $(\frac{n}{b})^{k-2}\binom{b}{k-2}$. (The worst case occurs when f_π corresponds to an equipartition of $[n]$ into b classes.) Using the fact stated earlier we see that the number of such complete $(b-k+2)$-partite graphs needed to decompose all the components of $C(k)$ is at least $\binom{n}{k-2}\frac{\log(n-k+2)}{\log(b-k+2)}$. Thus,

$$|\Pi|(\frac{n}{b})^{k-2}\binom{b}{k-2} \geq \binom{n}{k-2}\frac{\log(n-k+2)}{\log(b-k+2)}.$$

The theorem follows from this. □

This bound is weaker by a factor of $\frac{b-k+2}{b}$ when compared to the bound obtained by Körner.

Acknowledgment. I wish to thank Endre Szemerédi for valuable advice. I am grateful to Mark Kayll and Magnús Halldórsson for their comments on the previous version of this note. I would like to thank the referee for helpful suggestions.

References

[1] M. Fredman and J. Komlós. *On the Size of Separating Systems and Perfect Hash Functions,* SIAM J. Alg. Disc. Meth., 5 1985, pp. 61–68.

[2] J. Körner. *Fredman-Komlós Bound and Information Theory.* SIAM J. Alg. Disc. Meth., 7, 1986, pp. 560–570.

[3] I. Newman, P. Ragde, and A. Wigderson. *Perfect Hashing, Graph Entropy and Circuit Complexity* Proceedings of the fifth annual conference on Structure in Complexity Theory, 1990, pp. 91–99.

[4] J. Radhakrishnan. $\Sigma\Pi\Sigma$ *Threshold Formulas.* DIMACS Technical Report 90-73. December, 1990.

[5] M. Snir. *The Covering Problem of Complete Uniform Hypergraphs.* Discrete Mathematics 27, 1979, pp. 103–105.