Uniform Families and Count Matroids

Oleg Pikhurko*
Department of Pure Mathematics and Mathematical Statistics
Cambridge University, Cambridge CB3 0WB, UK
E-mail: `O.Pikhurko@dpmms.cam.ac.uk`

November 6, 2002

Abstract

Given an r-graph G on $[n]$, we are allowed to add consecutively new edges to it provided that every time a new r-graph with at least l edges and at most m vertices appears. Suppose we have been able to add all edges. What is the minimal number of edges in the original graph?

For all values of parameters, we present an example of G which we conjecture to be extremal and establish the validity of our conjecture for a range of parameters. Our proof utilises *count matroids* which is a new family of matroids naturally extending that of White and Whiteley.

We characterise, in certain cases, the extremal graphs. In particular, we answer a question by Erdős, Füredi and Tuza.

1 Uniform Families

An *r-graph* H is, as usual, a pair $(V(H), E(H))$ (*vertices* and *edges*), where $E(H)$ is a subset of $V(H)^{(r)} = \{A \subset V(H) : |A| = r\}$. The *size* of H is $e(H) = |E(H)|$ and the *order* is $v(H) = |V(H)|$. When isolated vertices do not matter, we usually identify H with $E(H)$; then, for example, $|H|$ stands for $e(H)$.

This research is motivated by the following problem which could be traced to Bollobás, see e.g. [2]. Given a family $\mathcal{F}$ of r-graphs, called *forbidden*, an r-graph G on $[n] = \{1, \ldots, n\}$ is called *weakly $\mathcal{F}$-saturated*, denoted $G \in \text{w-SAT}(n, \mathcal{F})$, if we can consecutively add all missing edges to G so that each time we add an edge a new forbidden subgraph appears. (Call the corresponding ordering of the edges of $\overline{G} = [n]^{(r)} \setminus G$, the *complement* of G, *$\mathcal{F}$-proper*.) What is the value of $\text{w-sat}(n, \mathcal{F})$, the minimal number of edges in the original graph?

*Supported by an External Research Studentship, Trinity College, Cambridge, UK.

Fix $l, m, r \in \mathbb{N}$ with $1 \le l \le \binom{m}{r}$. The *uniform family* $\mathcal{H} = \mathcal{H}_r(m, l)$ is the family of all r-graphs of order m and size l. By definition, $G \in \text{w-SAT}(n, \mathcal{H})$, $n \ge m$, if we can add the missing edges so that each creates a new subgraph with at most m vertices and at least l edges. For simplicity denote $\varepsilon(n, r, m, l) = \text{w-sat}(n, \mathcal{H}_r(m, l))$.

Uniform families were considered by Tuza [16] who made a conjecture on the value of $\varepsilon(n, r, r+1, l)$. Earlier results of Frankl [5] and Kalai [7, 8] (cf. also Lovász [12] and Alon [1]) imply the validity of the conjecture for $l = r+1$. Erdős, Füredi and Tuza [4] verified the conjecture for $l = 3$. Recently, the author [15] completely proved Tuza's conjecture and, besides, computed $\varepsilon(n, 2, m, l)$ for any n, m, l.

In this work we present, for all sets of parameters, a construction of G which we conjecture to be extremal. Our conjecture is in perfect accordance with the above results.

Clearly, our construction gives an upper bound on ε. To establish some lower bounds, we use the following observation of Kalai [8], see also Pikhurko [15]. Suppose we have a matroid $\mathcal{M}$ on $[n]^{(r)}$ such that every r-graph $F \in \mathcal{H}_r(m, l)$ with $V(F) \subset [n]$ is a circuit. Then any new edge, $\mathcal{H}$-properly added to G, must lie in the $\mathcal{M}$-closure of the existing edges. If $G \in \text{w-SAT}(n, \mathcal{H})$ then G must span $\mathcal{M}$ and therefore, $\varepsilon(n, r, m, l) \ge R(\mathcal{M})$, the rank of $\mathcal{M}$. Note that if we have equality, then every minimum weakly $\mathcal{H}$-saturated graph G forms a base of $\mathcal{M}$.

Thus, to prove some lower bounds we have to find some suitable $\mathcal{M}$. A family of matroids on $[n]^{(r)}$, called *count matroids*, was introduced by White and Whiteley [17], see also [19]. We generalise naturally the original definition to obtain a much wider family of matroids for which we preserve the same name. For example, our matroids admit many polynomials in n as the rank function while the previous definition is confined to linear functions only.

Applying count matroids we verify our conjecture for more sets of parameters although the question is still far from being resolved in the full generality. In certain cases, we characterise the sets of minimum weakly $\mathcal{H}$-saturated graphs. In particular, we answer a question by Erdős, Füredi and Tuza [4] who asked for a characterisation of the extremal graphs for $\mathcal{H}_r(r+1, 3)$.

We hope that count matroids will have other interesting applications; one (to scene analysis and geometry) is presented by Whiteley [18].

2 Construction

Let $n \ge m$, $1 \le l \le \binom{m}{r}$ and $\mathcal{H} = \mathcal{H}_r(m, l)$. We build, inductively on n, an example of a weakly $\mathcal{H}$-saturated graph $G_n = G(n, r, m, l)$ on $[n]$. If $n = m$ then, clearly, we can take for G_n any member of $\mathcal{H}_r(m, l-1)$. If $n > m$ then, inductively, choose any $G_{n-1} = G(n-1, r, m, l)$ and an $(r-1)$-graph $G' = G(n-1, r-1, m-1, l')$, where $l' = l - \binom{m-1}{r}$. (If $l \le \binom{m-1}{r} + 1$ then we take the empty graph for G'.) We let

$$G_n = G_{n-1} \cup \{E \cup \{n\} : E \in G'\}.$$

Let us show that G_n is indeed weakly $\mathcal{H}$-saturated. By the definition of G_{n-1}, we can add edges so that $[n-1]$ spans the complete r-graph. Then add edges $E_1 \cup \{n\}, \ldots, E_s \cup \{n\}$, where $(E_1, \ldots, E_s)$ is any $\mathcal{H}_{r-1}(m-1, l')$-proper ordering of the complement of G'. As each E_i creates a subgraph of size l' on some $(m-1)$-set $M \supset E_i$, $M \cup \{n\}$ spans at least $l' + \binom{m-1}{r} = l$ edges after $E_i \cup \{n\}$ has been added, which shows that $G_n \in \text{w-SAT}(n, \mathcal{H})$.

Conjecture 1 *For any $n, r, m, l \in \mathbb{N}$ with $m \le n$ and $1 \le l \le \binom{m}{r}$, $G(n, r, m, l)$ is a minimum weakly $\mathcal{H}_r(m, l)$-saturated graph.*

Remark. Generally, not all extremal graphs are given by our construction, cf. Theorem 7.

Let us compute the size of G_n. Given $l \ge 2$, define (uniquely) c and d so that

$$l = c + 1 + \sum_{j=0}^{d-1} \binom{m-j-1}{r-j}, \quad c \in [\tbinom{m-d-1}{r-d}], \quad d \in [0, r-1].$$

The definition of G_n implies, after some thought, the following formula for $e(G_n)$ which, alternatively, can be routinely checked by induction on n.

$$e(G_n) = \sum_{i=0}^{d} \left(c + \sum_{j=i}^{d-1} \binom{m-j-1}{r-j} \right) \binom{n-m+i-1}{i}, \quad n \ge m.$$

(We agree that $\binom{i}{0} = 1$, for *any* i.) For our purposes we have to find a representation $e(G_n) = \sum_{k=0}^{d} a_k \binom{n}{k}$. A substitution $\binom{n-m+i-1}{i} = \sum_{k=0}^{i} (-1)^{i-k} \binom{n}{k} \binom{m-k}{i-k}$ which is an instance of Vandermonde's convolution (see e.g. [6, p. 174]), implies

$$a_k = \sum_{i=k}^{d} (-1)^{i-k} \binom{m-k}{i-k} \left(c + \sum_{j=i}^{d-1} \binom{m-j-1}{r-j} \right).$$

Now, occasionally applying the identity $\sum_{i=0}^{t} (-1)^i \binom{j}{i} = (-1)^t \binom{j-1}{t}$, $t \ge 0$, we can find that $a_k = (-1)^{d-k} c \binom{m-k-1}{d-k} + (-1)^k s_k$, where

$$s_k = \sum_{j=k}^{d-1} \binom{m-j-1}{r-j} \sum_{i=k}^{j} (-1)^i \binom{m-k}{i-k} = (-1)^{d-1} \binom{m-k-1}{r-k} \binom{r-k-1}{d-k-1}.$$

Therefore, in summary,

$$e(G_n) = \sum_{k=0}^{d} (-1)^{d-k} \left(c \binom{m-k-1}{d-k} - \binom{m-k-1}{r-k} \binom{r-k-1}{d-k-1} \right) \binom{n}{k}.$$

3 Count Matroids

A function $\rho : X^{(<\infty)} \to \mathbb{R}$ (from finite subsets of X to the reals) is called *integral* if it is integer-valued, *increasing* if $\rho(A) \le \rho(B)$ whenever $A \subset B$, and *submodular* if

$$\rho(A \cup B) + \rho(A \cap B) \le \rho(A) + \rho(B), \quad A, B \in X^{(<\infty)}. \tag{1}$$

Given $\rho : 2^X \to \mathbb{R}$, we say that non-empty $A \subset X$ is *ρ-balanced* (or just *balanced* if ρ is understood) if $|A| \ge \rho(A) + 1$ but for every proper $B \subset A$ (that is $B \ne \emptyset$ and $B \ne A$) we have $|B| \le \rho(B)$.

Edmonds and Rota [3] observed the following.

Lemma 1 *For any integral, increasing and submodular function $\rho : 2^X \to \mathbb{R}$, the family of ρ-balanced sets satisfies the circuit axioms and therefore defines a matroid on X.* $\square$

The proof is easy and can be found for example in Oxley [13, Proposition 12.1.1].

We are interested in defining a matroid on $X = [n]^{(r)}$. (Then 2^X is identified with the set of r-graphs on $[n]$.) White and Whiteley [17], see also [19], introduced a family of *count matroids* on $[n]^{(r)}$ by defining $\rho(H) = a_1 |\cup_{E \in H} E| + a_0$, $H \subset [n]^{(r)}$, for some fixed a_1 and a_0.

We found it possible to generalise this construction in the following way. For $H \subset [n]^{(r)}$, we denote $p_i(H) = |\partial_i H|$, where

$$\partial_i H = \{D \in [n]^{(i)} : D \subset E \text{ for some } E \in H\}, \qquad i \in [0, r].$$

For example, $p_r(H) = e(H)$ and $p_1(H) = | \cup_{E \in H} E|$.

We consider *linear* functions, that is, functions defined by

$$L(H) = a_0 + \sum_{i=1}^{r-1} a_i p_i(H), \quad H \subset [n]^{(r)}, \tag{2}$$

for some constants $a_i \in \mathbb{R}$, $i \in [0, r-1]$.

Let us see when the function L on $X^{(<\infty)}$, $X = \mathbb{N}^{(r)}$, satisfies the above properties. Clearly, L is integral if and only if all coefficients are integers. Submodular and increasing linear functions are characterised by the following two lemmas which are of independent interest.

Lemma 2 *A linear function $L : X^{(<\infty)} \to \mathbb{R}$ is increasing if and only if*

$$\sum_{j=i}^{r-1} a_j \binom{r}{j} \ge 0, \qquad i \in [r-1]. \tag{3}$$

Proof. Suppose that L is increasing. Given $i \in [r-1]$, consider the r-graph $H = \{E \in [n]^{(r)} : |E \cap [r]| < i\}$, $n \geq 2r - i + 1$. We must have

$$L(H \cup \{[r]\}) - L(H) = \sum_{j=i}^{r-1} a_j \binom{r}{j} \geq 0,$$

which is exactly inequality (3).

On the other hand, suppose that L satisfies (3). Clearly, it is enough to show that, for any finite $H \subset X$ and $E \in X \setminus H$, we have $L(H) \leq L(H \cup \{E\})$. Let $C_i = \partial_i(H) \cap E^{(i)}$, $c_i = |C_i|/\binom{r}{i}$, $D_i = E^{(i)} \setminus \partial_i(H)$ and $d_i = |D_i|/\binom{r}{i}$, $i \in [r-1]$. Clearly, for any i and j, $1 \leq i < j \leq r-1$, the set system $D_i \cup C_j$ is an antichain in 2^E. By the LYM inequality $d_i \leq 1 - c_j = d_j$, that is,

$$0 \leq d_1 \leq \ldots \leq d_{r-1} \leq 1. \tag{4}$$

It is easy to check that

$$L(H \cup \{E\}) - L(H) = \sum_{i=1}^{r-1} a_i d_i \binom{r}{i}. \tag{5}$$

Consider the problem of minimising (5) given only the constraints (4). A moment's thought reveals that there exists $i \in [0, r-1]$ such that the extremum is achieved when $d_1 = \ldots = d_i = 0$ and $d_{i+1} = \ldots = d_{r-1} = 1$. But then (5) is non-negative by (3), so L is increasing. □

Lemma 3 *A linear function $L : X^{(<\infty)} \to \mathbb{R}$ is submodular if and only if $a_i \geq 0$, $i \in [r-1]$.*

Proof. The trivial consideration shows that, for any $i \in [r]$ and $H, G \subset [n]^{(r)}$, we have $p_i(H) + p_i(G) \geq p_i(H \cup G) + p_i(H \cap G)$. This implies (1) if every coefficient of L (except perhaps a_0) is non-negative.

On the other hand, suppose that L is submodular. Given any $i \in [r-1]$ consider the following set systems. Choose a 'large' m-set $Z \subset \mathbb{N}$ and $(r-i)$-sets D_Y and E_Y, indexed by $Y \in Z^{(i)}$, so that all $2\binom{m}{i} + 1$ selected sets are disjoint. Let

$$\begin{aligned} H &= \{D_Y \cup Y : Y \in Z^{(i)}\} \\ G &= \{E_Y \cup Y : Y \in Z^{(i)}\}. \end{aligned}$$

Clearly, we have $p_j(H \cap G) = 0$, $j \in [r-1]$, as $H \cap G = \emptyset$, and

$$p_j(H) = p_j(G) = \begin{cases} \binom{m}{i}\binom{r}{j}, & i < j \leq r-1, \\ \binom{m}{i}\left(\binom{r}{j} - \binom{i}{j}\right) + \binom{m}{j}, & 1 \leq j \leq i, \end{cases}$$

$$p_j(H \cup G) = \begin{cases} 2\binom{m}{i}\binom{r}{j}, & i < j \leq r-1, \\ 2\binom{m}{i}\left(\binom{r}{j} - \binom{i}{j}\right) + \binom{m}{j}, & 1 \leq j \leq i. \end{cases}$$

Routine calculations show that

$$L(H)+L(G)-L(H\cup G)-L(H\cap G)=a_i\binom{m}{i}+O(m^{i-1}),$$

which, by the submodularity of L, implies $a_i \geq 0$. $\square$

Thus we restrict our attention to integer coefficients satisfying

$$a_i \geq 0,\ i\in[r-1], \quad \text{and} \quad \sum_{j=0}^{r-1} a_j\binom{r}{j} \geq 1, \tag{6}$$

in which case, by Lemma 1, L defines a matroid $\mathcal{N}_L^n$ on $[n]^{(r)}$, $n \geq r$, which we still call a *count matroid.* The second condition in (6) excludes the degenerate case when already a single edge is dependent. Obviously, $\mathcal{N}_L^n$ is a *symmetric* matroid, that is, for any permutation σ of the vertex set $[n]$, $H\subset[n]^{(r)}$ is independent if and only if $\sigma'(H)$ is, where by σ' we denote the induced action on $[n]^{(r)}$. Clearly, the nested sequence $(\mathcal{N}_L^n)_{n\geq r}$ is compatible, so, usually, we do not specify n.

Actually, $\mathcal{N}_L$ admits an alternative definition if $a_0 \geq 0$. Let $X=[n]^{(r)}$ and let Y be the disjoint union of a_i copies of $[n]^{(i)}$, $i\in[0,r-1]$. Define the bipartite graph G on $X\cup Y$ by connecting $E\in X$ to all elements of Y corresponding to subsets of $E\in[n]^{(r)}$. (For example, every vertex in X has degree $\sum_{i=0}^{r-1} a_i\binom{r}{i}$.) It is easy to see that the *transversal matroid* of G, in which $H\subset X$ is independent if and only if H can be matched into Y, equals $\mathcal{N}_L^n$.

Any transversal matroid is representable over fields of every characteristics, see Piff and Welsh [14]; this applies to all count matroids with $a_0\geq 0$. It would be of interest to know whether $\mathcal{N}_L$ is representable for $a_0<0$.

Let us investigate the rank of $\mathcal{N}_L^n$.

Theorem 4 *Let L satisfy* (6). *Then the rank of $\mathcal{N}_L^n$ equals* $\min\left(\binom{n}{r}, L([n]^{(r)})\right)$.

Proof. We may assume that $\mathcal{N}=\mathcal{N}_L^n$ contains a non-trivial circuit for otherwise $R(\mathcal{N})=\binom{n}{r}\leq L([n]^{(r)})$ and our claim is true.

Let an r-graph G form a base for $\mathcal{N}$.

Claim 1. There exists an ordering of $\overline{G}=\{E_1,\dots,E_s\}$ such that

$$F_{[j-1]}\cap F_j \neq \emptyset, \quad j\in[2,s], \tag{7}$$

where F_i denotes the (unique and, by (6), non-empty) subgraph of G such that F_i+E_i is a circuit. (Also we denote $F_I=\cup_{i\in I}F_i$, $F+E=F\cup\{E\}$, etc.)

To prove the claim choose an arbitrary $E_1\in\overline{G}$ and, inductively, take for E_j any available edge satisfying (7). Suppose, on the contrary, that we are stuck after having chosen $E_1,\dots,E_{j-1}$, some $j\in[2,s]$. Let $G_1=F_{[j-1]}$ and $G_2=G\setminus G_1$. Both G_1 and G_2 are non-empty. Clearly, for any $E\in\overline{G}$ we must have either $F\subset G_1$ or

$F \subset G_2$ where $F + E$ is the circuit with $F \subset G$. Thus, if H_i is the closure of G_i, $i = 1, 2$, then $H_1 = G_1 + E_{[j-1]}$ and $H_2 = [n]^{(r)} \setminus H_1$.

Let C be any $\mathcal{N}$-circuit. We claim that C cannot intersect both H_1 and H_2. Suppose not. Let $E \in C \cap H_1$. As G_2 spans H_2, the rank of $(C \cap H_1) \cup G_2$ will not decrease if we remove E. Therefore, there is a circuit $C' \ni E$ such that $C' \subset (C \cap H_1) \cup G_2$. Likewise, fixing some $D \in C' \cap G_2 \neq \emptyset$, we obtain a circuit $C'' \subset G$ which contradicts the independence of G.

Note that if we replace C by the r-graph C' composed of the first $e(C)$ elements of $[n]^{(r)}$ in the colex order, then $p_i(C)$ would not increase by the Kruskal-Katona Theorem [10, 11], so $e(C') > L(C')$. If C' is not a circuit, take any proper subcircuit and repeat. The first two edges, $[r]$ and $[2, r+1]$, of the eventual circuit C' (which by (6) has size at least 2) share $r-1$ vertices and fall into the same half of $[n]^{(r)} = H_1 \cup H_2$. But every two edges can be connected by a sequence of edges such that any two neighbours share $r-1$ vertices. By the symmetry of $\mathcal{N}$, one of the halves must be empty, which is a contradiction proving Claim 1.

Choose an ordering guaranteed by Claim 1. Let us prove, by induction on j, the following.

Claim 2: $L(F_{[j]} + E_{[j]}) = L(F_{[j]}) = e(F_{[j]})$, $j \in [s]$.

First we note that, for every $i \in [s]$,

$$e(F_i) \leq L(F_i) \leq L(F_i + E_i) \leq e(F_i + E_i) - 1 = e(F_i),$$

which implies $L(F_i + E_i) = L(F_i) = e(F_i)$; in particular, our claim is true for $j = 1$. Now we argue as follows:

$$\begin{aligned} L(F_{[j]} + E_{[j]}) &\leq L(F_{[j-1]} + E_{[j-1]}) + L(F_j + E_j) - L(F_{[j-1]} \cap F_j) \\ &\leq e(F_{[j-1]}) + e(F_j) - e(F_{[j-1]} \cap F_j) \;=\; e(F_{[j]}). \end{aligned}$$

In the above transformations we use the submodularity of L, induction and the inequality $L(F_{[j-1]} \cap F_j) \geq e(F_{[j-1]} \cap F_j)$; the last inequality is valid because $F_{[j-1]} \cap F_j$ is independent and non-empty. (Actually, Claim 1 could be skipped if $a_0 \geq 0$.) Now, Claim 2 follows.

Clearly, $F_{[s]} = G$. Therefore, $L([n]^{(r)}) = L(G) = e(G) = R(\mathcal{N}_L^n)$. $\square$

Remark. Kalai [9] showed that, for any symmetric matroid $\mathcal{M}$ on $\mathbb{N}^{(r)}$, $R_{\mathcal{M}}([n]^{(r)})$ is a polynomial in n for all sufficiently large n. We call this polynomial the *growth polynomial* of $\mathcal{M}$. Kalai [9] also characterised all possible polynomials. Unfortunately, these are not confined to $L([n]^{(r)})$ with some L satisfying (6). (For example, the k-hyperconnectivity matroid on $\mathbb{N}^{(2)}$ introduced by Kalai [8] gives the polynomial $kn - \binom{k+1}{2}$.) It would be of interest to have a purely combinatorial construction (like that of a count matroid) producing every possible growth polynomial. (Matroids in [9] are constructed by means of multilinear algebra.)

4 Lower Bounds on $\varepsilon(n,r,m,l)$

Recall that the size of $G_n = G(n,r,m,l)$ is $\sum_{k=0}^{d} a_k \binom{n}{k}$, where

$$a_k = (-1)^{d-k}\left(c\binom{m-k-1}{d-k} - \binom{m-k-1}{r-k}\binom{r-k-1}{d-k-1}\right). \tag{8}$$

We define $L = \sum_{i=0}^{d} a_k p_k$, so that $L([n]^{(r)}) = e(G_n)$, the conjectured value. If L defines a matroid and every $F \in \mathcal{H}_r(m,l)$ is an $\mathcal{N}_L$-circuit then we can conclude that $\varepsilon(n,r,m,l) = e(G_n)$, which establishes the validity of our conjecture in this case. (Of course, this approach might work for other forbidden families.)

Now, the condition $a_k \geq 0$, $k \in [d]$, can be rewritten as

$$(-1)^{d-k}c \geq (-1)^{d-k}\frac{\binom{m-k-1}{r-k}\binom{r-k-1}{d-k-1}}{\binom{m-k-1}{d-k}} = (-1)^{d-k}\frac{d-k}{r-k}\binom{m-d-1}{r-d}.$$

The modulus of the latter expression is strictly decreasing with k, so, unfortunately, no suitable c would satisfy the conditions unless $d \leq 2$ and we have to confine ourselves to the three cases below.

4.1 $d = 0$

In this case the problem is trivial: it is easy to see directly that for $n \geq m \geq r \geq 1$ and $1 \leq l \leq \binom{m-1}{r} + 1$, $\varepsilon(n,r,m,l) = l-1$ and all extremal graphs are can be obtained by adding $n-m$ isolated vertices to some $H \in \mathcal{H}_r(m, l-1)$, which is exactly as our construction goes.

4.2 $d = 1$

Let $r \geq 2$ and $l = \binom{m-1}{r} + 1 + c$ with $1 \leq c \leq \binom{m-2}{r-1}$. By (8) we let $a_1 = c$ and $a_0 = \binom{m-1}{r} - c(m-1)$. The condition $1 \leq a_1 r + a_0$ implies that either $m = r+1$ and $c = 1$ or $m \geq r+2$ and

$$c \leq \min\left(\frac{\binom{m-1}{r} - 1}{m-r-1}, \binom{m-2}{r-1}\right) = \frac{\binom{m-1}{r} - 1}{m-r-1},$$

which we assume.

Let us show that every $F \in \mathcal{H}_r(m,l)$ is a circuit in $\mathcal{N}_L$. Obviously, $p_1(F) = m$, so $e(F) = L(F) + 1$ and F is not independent. Take any proper $F' \subset F$. If $p_1(F') = m$ then $L(F') = L(F) \geq e(F')$. If $p_1(F') \leq m-1$ then F' is independent by Theorem 4 as $L([m-1]^{(r)}) = \binom{m-1}{r}$. Hence F is a circuit and we obtain the following result.

Lemma 5 *Given r, m, l and n with $n \geq m > r \geq 2$ let $c = l - \binom{m-1}{r} - 1$. If $m > r+1$ and $1 \leq c < \frac{1}{m-r-1}\binom{m-1}{r}$ or if $m = r+1$ and $c = 1$ (when $l = 3$), then*

$$\varepsilon(n, r, m, l) = (l-1) + c(n-m). \quad \square$$

We can characterise extremal graphs in some cases by providing a combinatorial proof.

Lemma 6 *In addition to the assumptions of Lemma 5, assume that $m > r+1$ and $c < \frac{1}{m-1}\binom{m-1}{r}$. Then any minimum $G \in \text{w-SAT}(n, \mathcal{H})$ is given by our construction.*

Proof. Let $\overline{G} = \{E_1, \ldots, E_s\}$ be a proper ordering so that each E_i creates a forbidden subgraph on an m-set $M_i \subset [n]$ and let $L = a_1 p_1 + a_0$ be as above. We know that any $A \subset [n]$ spans at most $a_1|A| + a_0$ edges in G. (In fact, this is easy to see directly for otherwise we could replace these edges by a copy of $G(|A|, r, m, l)$, which would produce a smaller weakly $\mathcal{H}$-saturated graph.)

We prove by induction on i that, for any $i \in [s]$, $H_i \subset G$, the subgraph spanned by $M_{[i]} \subset [n]$, is given by our construction.

Clearly, this is the case for $i = 1$.

Let $i > 1$. We have to consider only the case when $k = |M_i \setminus M_{[i-1]}| \geq 1$. Of l edges of a forbidden subgraph F created by E_i, at most $\binom{m-k}{r}$ can belong to H_{i-1}, which shows that

$$|H_i \cup \{E_i\} \setminus H_{i-1}| \geq c + 1 + \binom{m-1}{r} - \binom{m-k}{r}.$$

It is routine to check that the last expression is strictly greater than $ck+1$ for $k \in [2, m]$. To prevent the contradiction $|H_i| > a_1|M_{[i]}| + a_0$, we must have $k = 1$ and $E_i \setminus M_{[i-1]} = \{x\}$ for some vertex x contained in exactly c edges of $F \cap G$. These edges (minus x) must lie within the $(m-1)$-set $M_{[i-1]} \cap M_i$, which is exactly what our construction says. $\square$

The value of $\varepsilon(n, r, r+1, 3)$ was computed by Erdős, Füredi and Tuza [4]. They asked if there is a characterisation of the extremal graphs. Our Lemma 6 does not cover this case, but we can provide a different proof of the lower bound which would give us the desired characterisation. Note that, up to isomorphism, $\mathcal{H}_r(r+1, 3)$ consists of a single graph.

Theorem 7 *For $\mathcal{H} = \mathcal{H}_r(r+1, 3)$ we have*

$$\text{w-sat}(n, \mathcal{H}) = n - r + 1, \qquad n \geq r. \tag{9}$$

Every extremal graph G can be obtained in the following way. Start with the set system G containing only one edge $[n]$. As long as possible, remove from G any edge E of size at least $r+1$, choose $A \in E^{(r-1)}$, partition $E \setminus A = X_1 \cup X_2$, $X_1, X_2 \neq \emptyset$, and add to G the edges $A \cup X_1$ and $A \cup X_2$.

Proof. Although we have already established (9), we have to provide a combinatorial proof of the lower bound. Let $G \in \text{w-SAT}(n, \mathcal{H})$. Note that every vertex in G is covered by at least one edge because, otherwise, the first edge added to G and containing this vertex cannot create a forbidden subgraph.

Let $E_1, \dots, E_j$ be the edges of G. With this sequence we do, step by step and as long as possible, the following operation. If some two sets have at least $r-1$ common points we merge them together, that is, replace them by their union (so the resulting system is no longer r-uniform).

We claim that we end up with the sequence containing a single member (which then must be equal to $V(G)$). Suppose not. Let $Y_1, \dots, Y_t$, $t \geq 2$, be the eventual family. Every two different resulting sets can have at most $r-2$ common points. Obviously, every edge of G lies within some Y_i. Let $E \in \overline{G}$ be the first edge added to G which does not lie entirely within some Y_i. (If for every $E \in [n]^{(r)}$ there is $Y_i \supset E$, then, considering chains of r-sets with overlaps of size $r-1$, we conclude that $Y_i = [n]$, some i.) The addition of E must have created $F \in \mathcal{H}$. The two other edges $E_1, E_2 \in E(F)$ either belong to G or were added before E and share $r-1$ vertices, so they lie each within some set Y_i. But then Y_i must contain $E \subset E_1 \cup E_2$ which is a contradiction. The claim is proved.

Now it is easy to prove by induction that in the above process every set of size m was a merger of at least $m-r+1$ edges of G. Trivially, it was the case for all initial sets which were precisely the edges of G. If we merge together 2 sets of sizes m_1 and m_2 made of $e_1 \geq m_1 - r + 1$ and $e_2 \geq m_2 - r + 1$ G-edges respectively, the resulting set has at most $m_1 + m_2 - r + 1$ vertices and $e_1 + e_2 \geq m_1 + m_2 - 2r + 2$ edges produced it, so the claim follows by induction.

If we have equality in (9), then in the merging procedure every two sets merged together have exactly $r-1$ common vertices, so every extremal graph can be obtained by reversing the merging process described in the statement of the theorem (of course in many different ways, generally).

We have to show that any anti-merging produces an extremal graph. Clearly, at the end we are left with r-subsets and we have exactly $n-r+1$ of these. To complete the theorem it is enough to show that a union of two complete r-graphs H_1 and H_2 of order at least r each with $|A| = |V(H_1) \cap V(H_2)| = r-1$, is weakly S-saturated. But this is easy: for $i = r-2, r-1, \dots$, add the missing edges which intersect A in exactly i points. $\Box$

Remark. Our construction of $G(n, r, r+1, 3)$ does not cover all cases as is demonstrated, for example, by $r = 3$, $n = 6$ and

$$G = \{\{1,2,3\},\ \{2,3,4\},\ \{4,5,6\},\ \{5,6,1\}\}.$$

4.3 $d = 2$

Assume $r \geq 3$ and $l = \binom{m-1}{r} + \binom{m-2}{r-1} + c + 1$ with $c \in [\binom{m-3}{r-2}]$. By (8), we let $a_2 = c$, $a_1 = -c(m-2) + \binom{m-2}{r-1}$ and $a_0 = c\binom{m-1}{2} - (r-1)\binom{m-1}{r}$.

Let us check when L satisfies (6). Of course, $a_2 \geq 1$. Next, the condition $a_1 \geq 0$ is, in our case, $c \leq \binom{m-2}{r-1}(m-2)^{-1}$. It is false for $m = r+1$, so assume $m \geq r+2$. The inequality $0 < a_2\binom{r}{2} + a_1 r + a_0$ reduces to

$$0 < c\binom{m-r-1}{2} + \left(r - \frac{(m-1)(r-1)}{r}\right)\binom{m-2}{r-1}. \tag{10}$$

Note that (10) is automatically true if $m = r+2$ (when the coefficient at c is zero), but then the condition $a_1 \geq 0$ implies $c = 1$. So, we conclude that L satisfies (6) if and only if either $m = r+2$ and $c = 1$ or $m \geq r+3$ and

$$\frac{((m-1)(r-1) - r^2)\binom{m-2}{r-1}}{r\binom{m-r-1}{2}} < c \leq \min\left(\frac{\binom{m-2}{r-1}}{m-2}, \binom{m-3}{r-2}\right) = \frac{\binom{m-2}{r-1}}{m-2}. \tag{11}$$

Let us check that any $F \in \mathcal{H}_r(m,l)$ is a circuit in $\mathcal{N}_L$. Clearly, every two vertices in F are covered by an edge for otherwise we would have at most $\binom{m}{r} - \binom{m-2}{r-2} < l$ edges in F. Therefore, $L(F) = L([m]^{(2)}) = l - 1 = e(F) - 1$ and we conclude that F is not $\mathcal{N}_L$-independent. On the contrary suppose that $L(H) < e(H)$ for some r-graph H on $[m]$ with at most $l-1$ edges. Clearly, we may assume that H is an initial segment of $[m]^{(r)}$ in the colex order.

Note that $L([m-1]^{(r)}) = \binom{m-1}{r}$ and, by Theorem 4, $[m-1]^{(r)}$ is independent. Therefore, H must have m vertices. Also, the 2-set $\{m-1, m\}$ cannot be covered by an H-edge, as then $e(H) \geq L([m]^{(r)}) + 1 \geq l$. Let H' be the $(r-1)$-graph on $[m-2]$ satisfying

$$H = [m-1]^{(r)} \cup \{D \cup \{m\} : D \in H'\}.$$

If we let $L' = a_2 p_1 + a_1$ then $L'([m-2]^{(r-1)}) = \binom{m-2}{r-1}$ and, by Theorem 4, $H' \subset [m-2]^{(r-1)}$ is independent in $\mathcal{N}_{L'}$ and $L'(H') \geq e(H')$.

Obviously, $p_2(H) = p_1(H') + \binom{m-1}{2}$. Therefore,

$$L(H) = L([m-1]^{(r)}) + L'(H') \geq \binom{m-1}{r} + e(H') = e(H),$$

which is the desired contradiction.

Theorem 8 *Assume that $r \geq 3$ and $l = \binom{m-1}{r} + \binom{m-2}{r-1} + c + 1$ such that either $m = r+2$ and $c = 1$ or $m \geq r+3$ and c satisfies* (11). *Then our conjecture is true.* $\square$

Unfortunately, we do not have any characterization of the extremal graphs.

Acknowledgements. The author is very grateful to Andrew Thomason for helpful comments.

References

[1] N. Alon. An extremal problem for sets with applications to graph theory. *J. Combin. Theory* (A), 40:82–89, 1985.

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

[3] J. Edmonds and G.-C. Rota. Submodular set functions (Abstract). *Waterloo Combinatorics Conference*, 1966.

[4] P. Erdős, Z. Füredi, and Z. Tuza. Saturated r-uniform hypergraphs. *Discrete Math.*, 98:95–104, 1991.

[5] P. Frankl. An extremal problem for two families of sets. *Europ. J. Combin.*, 3:125–127, 1982.

[6] R. L. Graham, D. E. Knuth, and O. Patashnik. *Concrete Mathematics: a Foundation for Computer Science.* Addison-Wesley Publ. Comp., 1989.

[7] G. Kalai. Weakly saturated graphs are rigid. *Ann. Discrete Math.*, 20:189–190, 1984.

[8] G. Kalai. Hyperconnectivity of graphs. *Graphs Combin.*, 1:65–79, 1985.

[9] G. Kalai. Symmetric matroids. *J. Combin. Theory* (B), 50:54–64, 1990.

[10] G. O. H. Katona. A theorem on finite sets. In *Theory of Graphs. Proc. Colloq. Tihany*, pages 187–207. Akademiai Kiado. Academic Press, New York, 1966.

[11] J. B. Kruskal. The number of simplices in a complex. In R. Bellman, editor, *Mathematical Optimization Techniques*, pages 251–278. Univ. California Press, Berkeley, 1963.

[12] L. Lovász. Flats in matroids and geometric graphs. In P. J. Cameron, editor, *Proc. 6th British Combin. Conf.*, pages 45–86. Academic Press, 1977.

[13] J. G. Oxley. *Matroid Theory.* Oxford Univ. Press, 1992.

[14] M. J. Piff and D. J. A. Welsh. On the vector representation of matroids. *J. Lond. Math. Soc.*, 2:284–288, 1970.

[15] O. Pikhurko. Weakly saturated hypergraphs and exterior algebra. To appear in *Combin. Prob. Computing*, 2001.

[16] Z. Tuza. Extremal problems on saturated graphs and hypergraphs. *Ars Combinatoria*, 25B:105–113, 1988.

[17] N. White and W. Whiteley. A class of matroids defined on graphs and hypergraphs by counting properties. Preprint, Math. Dept., Univ. Florida, Gainesville, FL 32611, 1984.

[18] W. Whiteley. A matroid on hypergraphs, with applications in scene analysis and geometry. *Discrete Comput. Geom.*, 4:75–95, 1989.

[19] W. Whiteley. Some matroids from discrete applied geometry. In J. E. Bonin, J. G. Oxley, and B. Servatius, editors, *Matroid Theory*, volume 197 of *Contemporary Mathematics*, pages 171–311. AMS, Providence RI, 1996.