Extremal Hypergraphs

Oleg Pikhurko

Department of Pure Mathematics and Mathematical Statistics
Cambridge University, UK
and
Trinity College, Cambridge, UK

E-mail: O.Pikhurko@dpmms.cam.ac.uk

October 11, 1999

Declaration

This is to certify that this dissertation is the outcome of my own work and no part was done in collaboration. To the best of my knowledge, all presented here results are original and new unless explicitly stated otherwise. No part of this dissertation has been submitted for a degree or a diploma at any other institution.

I want to express my extreme gratitude to Andrew Thomason, my research supervisor, for the very stimulating discussions, comments and encouragement which were of great help in the work. Without him this thesis could not have been possible.

I am thankful to Trinity College, Cambridge, for the financial support by an External Research Studentship and for the other kinds of support.

Oleg Pikhurko

Cambridge
December 21, 1999

Introduction

In this thesis we consider different extremal problems for set systems. The extremal (hyper-)graph theory has long been regarded as an important subject comprising a large number of various problems and methods.

Of course, we do not even try to present here all the features of the theory. Instead, we consider a few different facets such as saturated hypergraphs, weakly saturated hypergraphs, minimum chain decompositions, enumeration results for hypertrees, and size Ramsey numbers. We try to demonstrate different proof techniques in action and, indeed, the methods that we use are diverse: they include, for example, exterior algebra and probabilistic arguments.

Let us indicate how this work is organized. It is split into separate parts, each being a self-contained unit dealing with a particular feature. We tried as far as possible to ensure that each part can be read independently of the others. Please note that each part comes with its own introduction which can be consulted for further information.

Part I: Saturated Hypergraphs

Here we consider the notion of *saturation*. Let $\mathcal{F}$ be a family of forbidden *k-graphs*, that is, k-uniform set systems. A maximal k-graph G not containing any $F \in \mathcal{F}$ as a subgraph is called *$\mathcal{F}$-saturated*. We will be interested in $\mathrm{sat}(n, \mathcal{F})$, the minimal number of edges that an $\mathcal{F}$-saturated graph of order n can have. These types of questions were considered as early as the late 40s by Zykov [Zyk49], and by many other mathematicians henceforth.

However, there has been no good general upper bound on the sat-function. Tuza [Tuz86] (also an unpublished conjecture of Bollobás) conjectured that

$$\mathrm{sat}(n, F) = O(n^{k-1}), \qquad \text{for any fixed } k\text{-graph } F. \tag{1}$$

While the conjecture was proved for $k = 2$ by Kaszonyi and Tuza [KT86], and all particular examples confirmed its validity, it was not even known whether generally $\mathrm{sat}(n, F) = o(n^k)$ for $k \geq 3$. In Section 3 we verify this conjecture by showing that $\mathrm{sat}(n, \mathcal{F}) = O(n^{k-1})$ for all finite and certain infinite families $\mathcal{F}$ of k-graphs.

Different variations of the principle are presented in Section 4: we define the notion of saturation for different graph-like structures and investigate whether a form of (1) holds. While the technique of Section 3 extends to directed cycle-free graphs, ordered graphs, and layered graphs, we had to invent a new method in order to prove (1) for the class of k-row rectangular matrices.

In Subsection 4.4 we consider problems of the following type. Given a forbidden family, we say that a graph G *kills* an edge $E \in E(\overline{G})$ if the addition of E to G creates a forbidden subgraph. What is the maximal number of killed edges if G has a given order and size? We settle these problems for complete 2-graphs, which extends a theorem of Erdős, Hajnal and Moon [EHM64] who computed $\mathrm{sat}(n, K_m^2)$.

The sat-function is hard to handle: it lacks many natural regularity properties. For example, Kaszonyi and Tuza [KT86] showed that it is not monotone. In Section 5 we amplify their example: we construct, for any constant d, a 2-graph $F = F(d)$ such that $\mathrm{sat}(n, F) < \mathrm{sat}(n \pm 1, F) - d$ for a periodic series of values of n. Furthermore, we demonstrate a finite family $\mathcal{F}$ of 2-graphs for which the ratio $\mathrm{sat}(n, \mathcal{F})/n$ does not tend to a limit, which is rather unexpected and counterintuitive.

Specific instances of forbidden graphs are considered in Section 6.

We asymptotically compute $\mathrm{sat}(n, S_m^k)$, thus extending a result of Erdős, Füredi and Tuza [EFT91] who did the task for S_{k+1}^k. (The *generalized star* S_m^k is the k-graph on m vertices consisting of all k-tuples containing a given vertex.)

The *triangular family* $\mathcal{T}_k$ consists of all k-graphs of size 3 in which the symmetric difference of some two edges is contained in the third one. We prove that $\mathrm{sat}(n, \mathcal{T}_k) = n - O(\log n)$, $k \geq 3$, and $\mathrm{sat}(n, \mathcal{T}_3) = n - 2$.

We show that, for any K_m-saturated graph G, the number of edges spanned by the set $\{x \in V(G) : d(x) \leq a\}$ is at most $a^{2(m-2)a+o(ma)}$, a function not depending on $n = v(G)$. We deduce that G has at least $ln + O(\frac{n \log\log n}{\log n})$ edges, if the minimal degree of G is $l \geq m-1$. Another consequence is a sharper form of one result by Alon, Erdős, Holzman and Krivelevich [AEHK96, Theorem 2].

The following problem is, in fact, an instance of a sat-type problem. Suppose that we try to construct designs by adding, one by one and as long as possible, k-edges so that each t-set is covered by at most λ edges. What is the worst case, that is, how small the eventual system can be? We solve asymptotically this problem for $t = 2$ and establish some connections with Turán numbers for general t.

Part II: Weakly Saturated Hypergraphs

A notion related to that of saturation is *weak saturation* which we consider in Part II. A k-graph G is *weakly $\mathcal{F}$-saturated* if we can add one by one all missing edges to G so that every time at least one new forbidden subgraph appears; we are interested in w-sat$(n, \mathcal{F})$, the minimal size of a such graph G on n vertices.

These questions were first considered by Bollobás [Bol67c] who made a conjecture on complete graphs. The conjecture was verified by a number of people who computed w-sat(n, K_m^k): Frankl [Fra82], Kalai [Kal84, Kal85]; the result is implicit in Lovász [Lov77]; cf. also Alon [Alo85]. They all applied some form of dependence in order to derive the formula. This approach was most clearly formulated by Kalai [Kal85]: if we have a matroid $\mathcal{M}$ on $[n]^{(k)}$ such that any $F \in \mathcal{F}$ is a circuit, then w-sat$(n, \mathcal{F}) \geq R_{\mathcal{M}}([n]^{(k)})$, the rank of $\mathcal{M}$.

Usually, it is easy to construct a right example of minimum $G \in$ w-SAT$(n, \mathcal{F})$ for a given $\mathcal{F}$, but it is hard to prove that this G is indeed extremal. So, the above approach is helpful but it is not clear at all how to search for a suitable matroid $\mathcal{M}$.

Here we suggest two deterministic candidates for $\mathcal{M}$ to consider, provided we have an example of $G_n \in$ w-SAT$(n, \mathcal{F})$. For this purpose we utilize *gross* and *count* matroids which are defined in Section 8. The construction of a gross matroid was exploited by Kalai [Kal90], but for other purposes. Our *count matroids* form a new family of matroids, considerably and naturally extending the count matroid of White and Whiteley [WW84].

If one of our approaches works, then G is indeed extremal and we say that we have a *G-proof* or a *C-proof* respectively. Thus, we have two sufficient criteria for $G \in$ w-SAT$(n, \mathcal{F})$ to be minimal. Unfortunately, these criteria are not generally necessary, but using them (and the related *g/g'-proof* technique) we can prove the following results.

Given sequences of integers $\mathbf{s} = (s_1, \ldots, s_t)$ and $\mathbf{k} = (k_1, \ldots, k_t)$, the *pyramid* $P(\mathbf{s}; \mathbf{k})$ is the k-graph, $k = k_1 + \ldots + k_t$, with vertex set being the disjoint union $S_1 \cup \ldots \cup S_t$, $|S_i| = s_i$, and with the edge set consisting of those k-subsets which, for every $i \in [t]$, intersect $S_1 \cup \ldots \cup S_i$ in at least $k_1 + \ldots + k_i$ vertices. This is a rather general definition: as partial cases we obtain complete graphs and generalized stars.

In Subsection 10.1 we compute $\text{w-sat}(n, P(\mathbf{s}; \mathbf{k}))$ for *all feasible* values of parameters. A partial case of this result proves the conjecture by Tuza [Tuz88, Conjecture 7] that $\text{w-sat}(n, \mathcal{H}_k(k+1, l)) = \binom{n-k+l-2}{l-2}$, $n \geq k+1 \geq l \geq 2$. (The *uniform family* $\mathcal{H}_k(m, l)$ consists of all k-graphs with m vertices and l edges.)

In Subsection 10.2 we present some further results about weakly $\mathcal{H}_k(m, l)$-saturated graphs: we make a general conjecture and verify it for a number of parameters. In certain cases we characterize all extremal graphs, in particular answering a question by Erdős, Füredi and Tuza [EFT91] (who verified Tuza's conjecture for $l = 3$).

The *cone* $\text{cn}(F)$ of a k-graph F is obtained by adding an extra vertex x plus all $\binom{v(F)}{k-1}$ edges containing x. Our more general results of Section 11 imply that cones 'preserve' $G/g/g'$-proofs under certain covering conditions. This means that if we know the w-sat-function for certain graphs by applying a $G/g/g'$-proof, then we know it for the graphs obtained by the application of the cone operator. For example, for 2-graphs we can compute $\text{w-sat}(n, K_l + F)$, where for F we can take a star, an odd cycle, a path, a matching, and many other graphs.

In Section 12 we define *join*, another operation on graphs, and prove among other things that joins always preserve G/g-proofs. As a special case, we deduce the result of Alon [Alo85] who computed the w-sat-function for joins of complete hypergraphs.

Part III: Chain Decompositions

A *chain decomposition* of a poset $\mathcal{P}$ is a partition of $\mathcal{P}$ into disjoint *chains* (that is, linearly ordered subsets). Minimum chain decompositions have many applications and are extensively studied.

In this part we consider the minimal size of an *edge decomposition* which is a collection of skipless chains such that any pair $x \lessdot y$ (x is covered by y) belongs to exactly one chain. (A chain C is *skipless* if no element of $\mathcal{P} \setminus C$ can be inserted between some two elements of C.) It is easy to see that edge decompositions of $\mathcal{P}$ correspond to skipless chain decompositions of the *line poset* $L(\mathcal{P})$ whose vertex set is $\{(x,y) : x,y \in \mathcal{P},\ x \lessdot y\}$, and $(x,y) < (x',y')$ in $L(\mathcal{P})$ if $y \leq x'$ in $\mathcal{P}$.

In Section 14 we present a few min-max theorems. Our more general theorem implies that the minimal size of a skipless chain decomposition of $\mathcal{P}$ equals the maximal value of $|A| - |B|$ taken over all pairs of disjoint sets $A, B \subset \mathcal{P}$ such that any skipless chain containing two elements from A intersects B. Surprisingly enough, this fundamental theorem turned out to be a new result. Our proof utilizes the linear programming method of Dantzig and Hoffman [DH56]. It was considerably simplified by Graham Brightwell who replaced the linear programming argument by an easy application of Hall's theorem. We present both these proofs.

The minimal size of an edge decomposition of $\mathcal{P}$ can be deduced as a corollary, but we provide a short and direct proof.

Hence, our basic question is generally completely answered, but we can ask whether there is an edge decomposition with some extra properties. Of course, one can consider these problems for many different posets and impose many different restrictions. But as our theme is extremal set systems, we investigate $\mathcal{B}_n$, the poset of subsets of an n-set ordered by inclusion, and ask whether we can require that all chains are symmetric. (A skipless chain $A_1 \subset \ldots \subset A_k$ of $\mathcal{B}_n$ is *symmetric* if $|A_i| + |A_{k-i+1}| = n$, $1 \leq i \leq k$.) Note that any symmetric edge decomposition of $\mathcal{B}_n$ has the minimal size.

In fact, the general results of Anderson [And67] and Griggs [Gri77] imply the existence of a symmetric edge decomposition of $\mathcal{B}_n$. However, their proofs are non-constructive, so in Section 15 we provide an explicit construction.

Our decomposition has some extra properties and interesting applications, see Section 16. In brief, we give estimates of the number of antichains in $L(\mathcal{B}_n)$, construct a pair of orthogonal skipless chain decompositions of $L(\mathcal{B}_n)$, present some applications to storing and searching records in a database, and solve one numerical problem.

In Section 17 we characterize line posets in terms of forbidden configurations and point out which information determines and can be reconstructed from the line poset. (This resembles Beineke's [Bei68] characterization of line graphs.)

Part IV: Enumeration Results for Trees

Here we consider and enumerate different tree-like structures. Strictly speaking, such problems belong to enumerative, rather than to extremal, graph theory, but we include these results because we believe that the proofs are short and nice.

The notion of a tree and its different extensions to k-*graphs*, that is, k-uniform set systems, play an important role in discrete mathematics and computer science. We will dwell upon the following, rather general, definition suggested independently by Dewdney [Dew74] and Beineke and Pippert [BP77].

A k-graph is called a (k, m)-*tree* if it can be obtained from a single edge by consecutively adding edges so that every new edge contains $k - m$ new vertices while its remaining m vertices are covered by an already existing edge.

The problem of counting $(m + 1, m)$-trees which are known in the literature as m-*trees*, received great attention and was completely settled by Beineke and Pippert [BP69] and Moon [Moo69]. Later, different bijective proofs for m-trees appeared as well, see [RR70, Foa71, GI75, ES88, Che93].

Here we enumerate vertex labelled (k, m)-trees. We present two different proofs. The proof of Section 19 is inductive, that is, we write a recurrence relation for the number of trees and prove our formula by induction.

In Subsection 20.2 we exhibit an explicit bijection between the set of rooted vertex labelled trees of given size and a trivially simple set; it is based on the ideas of Foata [Foa71]. This method can be applied to enumerate other tree-like structures. For example, we enumerate vertex labelled k-gon trees. A k-*gon tree* is obtain from a k-gon (that is, a k-cycle) by consecutively adding k-gons along an existing edge, see e.g. [CL85, Whi88, Pen93, KT96]. In order not to repeat the same portions of proof twice, we present a more general result which includes both (k, m)-trees and k-gon trees as partial cases.

In Subsection 20.3 we present a bijection for *edge labelled* $(2, 1)$-trees, answering a question posed by Cameron [Cam95]. Unfortunately, we do not know any direct bijection enumerating edge labelled (k, m)-trees for general k, m.

Part V: Large Degrees in Subgraphs

Erdős [Erd81], see also [Chu97, Erd99], conjectured that for $n \geq 3$ any graph with fewer than $\binom{2n+1}{2} - \binom{n}{2} = \frac{3n(n+1)}{2}$ edges is a union of a bipartite graph and a graph with maximum degree less than n. All research carried in this part is motivated by this conjecture which is disproved here.

The conjectured value arises from the consideration of $P_{n+1,n} = K_{n+1} + E_n$ which does not admit the above representation. In fact, this graph has a stronger property, namely $P_{n+1,n} \to (K_{1,n}, K_3)$: for any blue-red colouring of the edge set of $P_{n+1,n}$ we necessarily have either a blue star $K_{1,n}$ or a red triangle. Thus, if Erdős' conjecture were true, it would give the same value for the *size Ramsey number* $\hat{r}(K_{1,n}, K_3) = \min\{e(G) : G \to (K_{1,n}, K_3)\}$. Apparently, the computation of $\hat{r}(K_{1,n}, K_3)$ was the original motivation for the conjecture.

In Section 22 we show, however, that

$$\hat{r}(K_{1,n}, K_3) < n^2 + \sqrt{2}\, n^{3/2} + n, \quad n \geq 1,$$

by demonstrating an explicit construction. This disproves Erdős' conjecture which, in fact, fails for all $n \geq 5$. On the other hand, we prove that any graph with $n^2 + (0.577 + o(1))\, n^{3/2}$ edges is a union of a bipartite graph and a graph with maximum degree less than n, which of course implies that this number is a lower bound for $\hat{r}(K_{1,n}, K_3)$.

There were different attempts to prove the conjecture, by different mathematicians, which resulted in new interesting directions of research.

For example, as reported in [Erd99], Erdős and Faudree [EF99] consider the minimal size of a graph G such that if G is a union of two graphs, one having maximal degree less than n, then the other contains all odd cycles C_m with $3 \leq m \leq n-3$. In Subsection 22.3 we demonstrate a graph G of size $(1+\varepsilon)n^2$, for any given constant $\varepsilon > 0$, such that, for any blue-red colouring of G without a blue $K_{1,n}$, we have red cycles of all lengths (odd and even) between 3 and cn, where $c = c(\varepsilon) > 0$ does not depend on n.

The following problem, which was introduced by Erdős, Reid, Schelp and Staton [ERSS96], is also motivated by Erdős' conjecture.

For positive integers n, k, j with $k \geq j$, let $\mathcal{M}(n,k,j)$ consist of all graphs G of order $n+k$ such that every $(n+j)$-subset of $V(G)$ spans a graph with maximum degree at least n. The question is to compute

$$m(n,k,j) = \min\{e(G) : G \in \mathcal{M}(n,k,j)\}.$$

Erdős et al [ERSS96, Conjecture 1] conjectured that, for $n \geq k \geq j \geq 1$ and $n \geq 3$, we have

$$m(n,k,j) = (k-j+1)n + \binom{k-j+1}{2}. \tag{2}$$

This value arises from the consideration of $P_{k-j+1,n} \sqcup E_{j-1}$. Erdős et al [ERSS96, Theorem 3] proved that (2) is true if $j = 1$ or if $j \geq 2$ and

$$n \geq \max\left(j(k-j),\ \tbinom{k-j+2}{2}\right). \tag{3}$$

In Section 23 we demonstrate a constructive counterexample to (2) for $n \leq (j-2)(k-j)$. On the other hand, we show that (2) is true if

$$n \geq \max\left(\left(j+\tfrac{1}{2}\right)(k-j) + \tfrac{j+k}{4j-2}, 14\right),$$

which improves (3) for $j \lessapprox k/3$. This shows that $j(k-j)$ is roughly the threshold on n when the obvious construction leading to (2) fails to be extremal. Some other constructions are presented.

In Section 24 we consider the following related problem. Let $\mathcal{B}(n,m)$ consist of all graphs such that for any partition $V(G) = A \cup B$ either $\Delta(G[A]) \geq n$ or $\Delta(G[B]) \geq m$ (or both). We are interested in the *bisplit function*

$$b(n,m) = \min\{e(G) : G \in \mathcal{B}(n,m)\}.$$

Clearly, $b(n,n)$ is precisely the function investigated in Erdős' conjecture, which was the original motivation for introducing the 'off-diagonal' numbers $b(n,m)$.

We compute this function asymptotically when $m = \min(n,m)$ is large:

$$b(n,m) = 2nm - m^2 + o(m)n.$$

In the extreme case, when $m \geq 1$ is fixed, we can prove only that the numbers $b(n,m)$, $n \in \mathbb{N}$, lie between two functions linear in n with slopes $2m+1$ and $2m + \sqrt{2m} + \frac{5}{2}$.

We prove that $b(n,1) = 4n-2$ for $n \geq 8$ (and characterize all extremal graphs) and that $b(n,2) = 6n + O(1)$. As the reader will see, the proofs are rather lengthy and require consideration of many cases. This indicates that the computation of $\lim_{n\to\infty} b(n,m)/n$ for any fixed m (if the limit exists) is perhaps a hard task.

Notation

Let us indicate some notation that we use. The relation $A \subset B$ does not exclude $A = B$; the strict inclusion is denoted as $A \subsetneq B$. Any unfamiliar term (e.g. pyramid) should be identifiable via the index.

$[m,n]$	$=$	$\{m, m+1, \ldots, n\}$; $[n] = \{1, 2, \ldots, n\}$
$A^{(r)}$	$=$	$\{B \subset A : \lvert B\rvert = r\}$
$\mathbb{R}/\mathbb{Q}/\mathbb{Z}/\mathbb{N}$		the sets of reals/rationals/integers/positive integers
$f = \Theta(g)$	$\Leftrightarrow$	$\exists c_1, c_2 > 0 \ \ \exists n_0 \ \ \forall n \geq n_0 \quad c_1 g(n) \leq \lvert f(n)\rvert \leq c_2 g(n)$
$f = O(g)$	$\Leftrightarrow$	$\exists c > 0 \ \ \exists n_0 \ \ \forall n \geq n_0 \quad \lvert f(n)\rvert \leq cg(n)$
$f = o(g)$	$\Leftrightarrow$	$\forall c > 0 \ \ \exists n_0 \ \ \forall n \geq n_0 \quad \lvert f(n)\rvert \leq cg(n)$
s_A	$=$	$\sum_{i \in A} s_i$, given reals $s_1, \ldots, s_n$ and $A \subset [n]$
B_A	$=$	$\cup_{i \in A} B_i$, given sets $B_1, \ldots, B_n$ and $A \subset [n]$
$V(G)$		the vertex set of G
$v(G) = \lvert V(G)\rvert$		the order of G
$E(G)$		the edge set of G
$e(G) = \lvert E(G)\rvert$		the size of G
$\overline{G}$		the complement of G
$G[A]$		the subgraph induced by $A \subset V(G)$
$\alpha(G)$		the independence number of G
$d(x)$	$=$	$\lvert\{E \in E(G) : E \ni x\}\rvert$, $x \in V(G)$
$\Delta(G)/\delta(G)$		the maximal/minimal degree of G
$\Gamma_A(x)$	$=$	$\{y \in A : \{x, y\} \in E(G)\}$, 2-graph G; $\Gamma(x) = \Gamma_{V(G)}(x)$
$d_A(x)$	$=$	$\lvert\Gamma_A(x)\rvert$, $x \in V(G)$, $A \subset V(G)$, 2-graph G
mG		m disjoint copies of G
C_m		the m-cycle
E_m		the empty graph of order m
$K^k(A)$		the complete k-graph on a set A
K_m^k		the complete k-graph of order m; $K_m = K_m^2$
$K_{m,n}$		the complete bipartite graph
$P_{m,n}$	$=$	$K_m + E_n$
$P(\mathbf{s}; \mathbf{k})$		the pyramid
S_m^k		the k-star of order m

Contents

Part I

Saturated Hypergraphs

1 Introduction

1.1 Discussion

Many combinatorial structures (especially graphs) have proved to be very useful in other branches of human knowledge where discrete models play more and more important role with the advance of computers. A fairly typical problem is, given a class $\mathcal{C}$ of allowed graphs (for example, those whose structure is compatible with the requirements of the used discrete model), to minimize/maximize a certain parameter.

In many natural cases, $\mathcal{C}$ can be described by naming a family $\mathcal{F}$ of forbidden subgraphs so that a graph belongs to $\mathcal{C}$ if and only if it is *$\mathcal{F}$-free*, that is, if it does not contain any $F \in \mathcal{F}$ as a subgraph. In this case, $\mathcal{C}$ can be also specified by listing the family $\mathrm{SAT}(\mathcal{F})$ of all *$\mathcal{F}$-saturated* graphs, that is, maximal $\mathcal{F}$-free graphs; clearly,

$$\mathcal{C} = \{H : H \subset G \text{ for some } G \in \mathrm{SAT}(\mathcal{F})\}$$

and, instead of considering the whole of $\mathcal{C}$, we can restrict ourselves only to $\mathrm{SAT}(\mathcal{F})$, especially that many extremal parameters of $\mathcal{C}$ can be more quickly determined from $\mathrm{SAT}(\mathcal{F})$.

Two related families are m-$\mathrm{SAT}(\mathcal{F})$ and w-$\mathrm{SAT}(\mathcal{F})$: $G \in$ m-$\mathrm{SAT}(n,\mathcal{F})$ if the addition any new edge to G creates at least one new forbidden subgraph (then we call G *monotonically $\mathcal{F}$-saturated*); $G \in$ w-$\mathrm{SAT}(n,\mathcal{F})$ if we can add all missing edges, one by one in some order, so that every edge creates a new forbidden subgraph (then we call G *weakly $\mathcal{F}$-saturated*). Note that we do not require here that G is $\mathcal{F}$-free.

For example, the Turán-type problem studies the maximal size of an $\mathcal{F}$-free graph of a given order. This is clearly equal to the maximal size of an $\mathcal{F}$-saturated graph of a given order.

In Parts I and II we consider the sat-type problems which ask about the minimal size of a (weakly/monotonically) $\mathcal{F}$-saturated graph of a given order.

The Turán-problem and the sat-problem happen to be rather different in nature. The former is perhaps more important in applications although many real life situations lead to sat-type questions.

For example, monotonically K_3-saturated graphs are precisely diameter-2 graphs. This problem has the following interpretation: there are n airports; we can connect some pairs by a flight and we want to assure the possibility to fly from each airport to any other one by changing the plane at most once. Clearly, the minimal number of connections is $n-1$ and this is achieved if and only if one airport is connected to every other. (This may be not a perfect solution and we may impose some extra conditions: e.g. some restriction on the maximal degree, etc.) If we weaken the requirement by allowing any number of flight changes, then we obtain weakly K_3-saturated (i.e. connected) graphs and the minimal size is again $n - 1$ but we have many extremal graphs.

In this part (and Part II) we try to present a unified treatment of these, sat-type, questions. The above definitions are applied to *k-graphs* (k-uniform set systems) which are the main object of our consideration. Also, we present different variations of the principle and make a few excursions into some related areas (for example, the forbidden submatrix problem). Section 2 briefly surveys known results on the topic including those proved here. But before we proceed, let us give all necessary definitions.

1.2 Definitions

Let $\mathcal{F}$ be a family *k-graphs* (that is, k-uniform set systems) which are usually referred to as *forbidden*. A k-graph G is called *$\mathcal{F}$-admissible* (or *$\mathcal{F}$-free*) if it does not contain any $F \in \mathcal{F}$ as a subgraph.

We say that G is *$\mathcal{F}$-saturated*, denoted $G \in \mathrm{SAT}(n, \mathcal{F})$, if it is a maximal $\mathcal{F}$-free k-graph with n vertices. We are mainly interested in

$$\mathrm{sat}(n, \mathcal{F}) = \min\{e(G) : G \in \mathrm{SAT}(n, \mathcal{F})\}, \tag{4}$$

the minimal number of edges in an $\mathcal{F}$-saturated graph of order n.

The following auxiliary notion is helpful: G is called *monotonically* (or *strongly*) $\mathcal{F}$-saturated, denoted $G \in \text{m-SAT}(n, \mathcal{F})$, $n = v(G)$, if the addition of any new edge to G creates at least one extra F-subgraph, some $F \in \mathcal{F}$. Note that we do not require that G is $\mathcal{F}$-admissible.

Clearly, $\mathrm{SAT}(n,\mathcal{F}) = \{G \in \text{m-SAT}(n,\mathcal{F}) : G \text{ is } \mathcal{F}\text{-free}\}$, so $\mathrm{sat}(n,\mathcal{F}) \geq \text{m-sat}(n,\mathcal{F})$, where

$$\text{m-sat}(n,\mathcal{F}) = \min\{e(G) : G \in \text{m-SAT}(n,\mathcal{F})\}.$$

For a graph F, we denote $\mathrm{SAT}(n,F) = \mathrm{SAT}(n,\{F\})$, etc.

2 Survey

Here is a brief but comprehensive (to the best of the author's knowledge) survey of known results related to (strong) saturation. Also, we indicate all interesting results proved in this part.

2.1 General Families

Not much is known about $\mathrm{sat}(n,\mathcal{F})$ for a general $\mathcal{F}$. Kászonyi and Tuza [KT86] showed that, for any family $\mathcal{F}$ of 2-graphs, including all infinite families, we have $\mathrm{sat}(n,\mathcal{F}) = O(n)$. Tuza [Tuz92] showed that, for any fixed k-graph F,

$$\text{m-sat}(n,F) = \Theta(n^{d(F)}). \tag{5}$$

Here $d(F) \in [0, k-1]$ is what Tuza calls the *local density* of F:

$$d(F) = \min\{d(E) : E \in E(F)\}, \tag{6}$$

where the *density* $d(E)$ of an F-edge E is $\max\{|E \cap E'| : E' \in E(F)\ E' \neq E\}$.

Clearly, in terms of constructive upper bounds, SAT is more restrictive than m-SAT. Thus, it is not surprising that, up to now, there were no good upper bounds on $\mathrm{sat}(n,F)$ for a general k-graph F. Tuza [Tuz86, Tuz88] (also an unpublished conjecture of Bollobás) conjectured that, for any fixed k-graph F, $\mathrm{sat}(n,F) = O(n^{k-1})$.

In Section 3 we show that

$$\mathrm{sat}(n,\mathcal{F}) = O(n^{k-1}) \tag{7}$$

for all finite and certain infinite families $\mathcal{F}$, which, of course, proves this conjecture. Our proof is constructive.

In Section 4 we try to extend the notion of saturation to different structures connected to hypergraphs and every time we ask whether the analogue of estimate (7) is valid. Although the estimate is not true for simple directed graphs, we show that (7) is valid for all finite families of cycle-free directed k-graphs and for ordered k-graphs. Furthermore, the estimate $\mathrm{sat}(n, \mathcal{F}) = O(n)$ is true for *any* family $\mathcal{F}$ of cycle-free or ordered 2-graphs.

In Subsection 4.2 we consider similar question for structures that we call *layered graphs* and show that a form of (7) holds here. Also, we show that, for the class of layered $(1, 1)$-graphs (that is, bipartite graphs), the size of any minimum $\mathcal{F}$-saturated graph is bounded by a linear function of its order for *any* forbidden family $\mathcal{F}$.

In Subsection 4.3 we consider the sat-type problems for the class of rectangular matrices, for which the dual (Turán-type) problems are well studied. We show that for any family $\mathcal{F}$ of forbidden k-row matrices $\mathrm{sat}(n, \mathcal{F}) = O(n^{k-1})$.

Although the notion of saturation was considered as early as the late 40s by Zykov [Zyk49], the theory does not seem to be well developed. This might be the case because minimum saturated graphs are hard to handle. For example, as demonstrated by Kászonyi and Tuza [KT86], the sat-function lacks many natural regularity properties; in Section 5 we provide further examples.

Answering a question by Tuza [Tuz92] we exhibit an example of connected 2-graphs $H \subset F$ of the same order such that $\mathrm{sat}(n, H) > \mathrm{sat}(n, F)$ for all large n. (Of course, it is 'natural' to expect the converse inequality.)

Among other things, we demonstrate, for any fixed $d > 0$, a 2-graph $F = F(d)$ such that

$$\mathrm{sat}(n, F) < \mathrm{sat}(n \pm 1, F) - d,$$

for a periodic series of values of n.

Tuza [Tuz88] conjectured that, for any 2-graph F, the limit $\lim \mathrm{sat}(n, F)/n$ exists. Of course, a number of similar questions arise for k-graphs as well. Unfortunately, there is not much progress in this direction.

Truszczynski and Tuza [TT91], characterized those 2-graphs F for which $c = \lim \mathrm{sat}(n, F)/n$ exists and is smaller than 1; then, in fact, $c = 1 - 1/p$, $p \in \mathbb{N}$.

In Section 5 we demonstrate a finite family $\mathcal{F}$ of 2-graphs for which the limit

$\lim \mathrm{sat}(n, \mathcal{F})/n$ does not exist.

In the literature, there are many different variations on the topic; one possibility is to consider minimum saturated graphs (most frequently K_m^2-saturated) with some extra restrictions, for example, on degrees (Hajnal [Haj65], Hanson and Seyffarth [HS84], Duffus and Hanson [DH86], Erdős and Holzman [EH94], Füredi and Seress [FS94], Alon et al [AEHK96]), chromatic number (Hanson and Toft [HT91]), etc. Hanson and Toft [HT87] consider edge-coloured saturated graphs.

2.2 Particular Cases

Erdős, Hajnal and Moon [EHM64] via an inductive argument and contractions computed the sat-function for all complete 2-graphs. Bollobás [Bol65] introduced the powerful weight method and proved that

$$\mathrm{sat}(n, K_m^k) = \binom{n}{k} - \binom{n-m+k}{k}, \quad n \geq m > k. \tag{8}$$

The cases of equality were characterized in both papers.

We show that, for any K_m-saturated graph G, the number of edges spanned by the set $\{x \in V(G) : d(x) \leq a\}$ is bounded by $a^{2(m-2)a+o(ma)}$, a function of a and m only. We deduce that G has at least $ln + O(\frac{n \log\log n}{\log n})$ edges, $n = v(G)$, if the minimal degree of G is $l \geq m-1$. Another consequence is a sharper form of one result by Alon, Erdős, Holzman and Krivelevich [AEHK96, Theorem 2]. Please refer to Subsection 6.4 for details.

The *star* S_m^k has m vertices and consists of k-tuples containing a fixed vertex. The *uniform family* $\mathcal{H}_k(m, l)$ consists of all k-graphs of order m and size l. Erdős, Füredi and Tuza [EFT91] determined the exact sat-values for $\mathcal{H}_3(6, 3)$ and $\mathcal{H}_3(4, 3) = S_4^3$ and described the cases of equality. Also, they found asymptotic values for $\mathcal{H}_k(k+1, k) = S_{k+1}^k$. In Subsection 6.1 we extend the last result by computing asymptotically $\mathrm{sat}(n, S_m^r)$ for all possible r and m.

In Subsection 6.2 we define a t-(v, k, λ)-*sub-design* G as a maximal k-graph of order n such that no t-set is covered by more than λ edges. (Sub-designs naturally arise when we try to construct designs by consecutively adding edges as long as possible.) If we let $\mathcal{D} = \mathcal{D}(\lambda, k, t)$ be the family of all k-graphs with $\lambda + 1$ edges sharing at least t common vertices then $\mathrm{SAT}(n, \mathcal{D})$ is the family of all sub-designs of order n. We compute exactly $\mathrm{sat}(n, \mathcal{D}(\lambda, k, t))$ for $t = 1$ and any λ, k, n (except for a few small values of n) and (asymptotically) for $t = 2$ and any fixed λ, k. In the general case $t \geq 3$ we deduce some lower bounds and establish connections with the Turán problem for complete hypergraphs.

In Subsection 6.3 we forbid 3 edges such that the symmetric difference of some two edges is contained in the third one and compute asymptotically the corresponding sat-function. (For 3-graphs, we find the exact value.)

Erdős and Gallai [EG61] showed that mK_3^2 is the (unique) minimum graph in $\mathrm{SAT}(n, mK_2^2)$ for $n \geq 3m$. (By mF we denote the union m disjoint copies of F.) The case of mK_k^k, $k \geq 3$, is harder. Many authors present different lower and upper bounds on $\mathrm{sat}(n, 2K_k^k)$ for specific k. The best known general bounds seem to be $\mathrm{sat}(n, 2K_k^k) \leq k^5$, $k \geq 1$, by Blokhuis [Blo87], and $\mathrm{sat}(n, 2K_k^k) \geq 3k$, $k \geq 4$, by Dow et al [DDFL85].

Wessel [Wes66, Wes67] and Bollobás [Bol67b, Bol67a] computed independently the sat-function and characterized extremal graphs for all complete bipartite graphs in the class of bipartite, that is, $(1,1)$-layered, graphs.

Concerning 2-graphs, Kászonyi and Tuza [KT86] found the complete answer for all paths and stars. The situation for cycles looks rather complicated. Of course, the case $C_3 = K_3^2$ is known. Ollman [Oll72] proved that $\mathrm{sat}(n, C_4) = \lfloor (3n-5)/2 \rfloor$ and all extremal graphs were described by Tuza [Tuz89]. According to a recent paper by Barefoot et al [BCE+96], for every $k \geq 5$, we know the exact values of $\mathrm{sat}(n, C_k)$ only for finitely many values of n although some general bounds are available.

A result of Bondy [Bon72b] implies that

$$\mathrm{sat}(n, C_n) \geq \lceil 3n/2 \rceil. \tag{9}$$

There was a great amount of work invested in computing this function exactly (Isaacs [Isa75], Clark et al [CE83, CCES86, CES92]) until the computation was completely finished by Xiaohui et al [XWCY97] (with final touches made by computer search). In fact, estimate (9) is sharp for all even $n \geq 20$ and all odd $n \geq 17$.

Füredi et al [FHPZ98] considered digraphs and showed that $\mathrm{sat}(n, \overrightarrow{C_3}) = (1 + o(1)) n \log_2 n$. (Here $\overrightarrow{C_3}$ denotes the directed 3-cycle.)

In Subsection 4.4 we investigate the maximal number of edges which cannot be $\mathcal{F}$-freely added to G, given $v(G)$ and $e(G)$. We settle this problem (with a description of all extremal graphs) for complete 2-graphs, which extends the already mentioned result of Erdős, Hajnal and Moon [EHM64] who computed $\mathrm{sat}(n, K_m^2)$.

3 Construction

Here we demonstrate some constructive upper bounds on $\mathrm{sat}(n,\mathcal{F})$ for a general family $\mathcal{F}$ which, in particular, imply the conjecture of Tuza [Tuz86] (also conjectured by Bollobás, unpublished) that, for any k-graph F,

$$\mathrm{sat}(n, F) = O(n^{k-1}). \tag{10}$$

Note that we cannot replace $k-1$ by a smaller exponent in (10) if we want the estimate to be valid for every k-graph F; this follows, for example, from formula (8).

Kászonyi and Tuza [KT86] proved that $\mathrm{sat}(n,\mathcal{F}) = O(n)$, for any family $\mathcal{F}$ of forbidden 2-graphs, including infinite families; this verifies (10) for $k = 2$. However, there has been no progress in proving (10) for $k \geq 3$ and the conjecture is mentioned in a few different papers, e.g. in [Tuz88, EFT91, Tuz92, Fra95]. Also, the importance of estimate (10) might be indicated by the fact that Bollobás [Bol95], in his authoritative survey of the whole of extremal graph theory, gives two different proofs of $\mathrm{sat}(n,\mathcal{F}) = O(n)$ for 2-graphs.

Let us present some general construction of $H \in \mathrm{sat}(n,\mathcal{F})$ which implies (10); this result appears in [Pik99d].

For a k-graph H, we say that $A \subset V(H)$ is *independent* if it does not span an edge in H, that is, $A^{(k)} \cap E(H) = \emptyset$.

Theorem 1 *Let $\mathcal{F}$ be a family of k-graphs. Suppose that there is $s \in \mathbb{N}$ such that no $F \in \mathcal{F}$ contains an independent set $A \subset V(F)$ of order $s+1$ which can be covered by a union of F-edges sharing a common vertex outside A. Then, for any n,*

$$\mathrm{sat}(n,\mathcal{F}) < \left(s' - s + 2^{k-1}(s-1)\right)\binom{n}{k-1}, \tag{11}$$

where $s' = \min\{v(F) : F \in \mathcal{F}\}$.

Proof. It is enough to construct a graph $H \in \mathrm{SAT}(n, \mathcal{F})$ whose size does not exceed the stated bound. Our construction will be by means of an algorithm.

Our algorithm works in the following way. Let us agree that the vertex set is $X = [n]$ with the usual ordering. Given $x \in X$ and $B \subset X$, we write $B < x$ if every vertex in B is smaller than x. By $U_x = \{y \in X : y > x\}$ we denote the *upper shadow* of x and in the obvious way we define the *lower shadow* L_x. If $|B| \leq k$, say B consists of elements $b_1 < \ldots < b_i$, $i \leq k$, then we define its *tail*

$$\mathcal{T}_B = \{\{b_1, \ldots, b_i, x_{i+1}, \ldots, x_k\} : b_i < x_{i+1} < \ldots < x_k\} \subset X^{(k)}. \qquad (12)$$

We construct an $\mathcal{F}$-saturated graph H by starting with the empty hypergraph H on X and adding to H one by one certain families of edges until we obtain $H \in \mathrm{SAT}(n, \mathcal{F})$.

The algorithm is rather simple. We take, one by one in order, the vertices of X. For every vertex x, we consider all of the i-subsets of L_x, beginning with $i = 0$ and increasing i until $i = k - 1$. For every such subset $A < x$, we consider $\mathcal{T}_B$, $B = A \cup \{x\}$, which is, by the definition, the family of k-subsets having B as an initial segment. If at this moment $\mathcal{T}_B \not\subset E(H)$ and the addition of $\mathcal{T}_B$ to the edge set of H does not create any forbidden subgraph, we add $\mathcal{T}_B$ to H. This is a crucial feature of the algorithm: for every x and A we either add *all* of $\mathcal{T}_B$ or we add *nothing.*

Another important detail is the order of the steps. The outermost cycle has x *increasing* from 1 to n. The next cycle runs for i *increasing* from 0 to $k - 1$. In the innermost cycle we consider all i-subsets of L_x and here we are free to choose them in any order, but for uniformity let us agree that we use here the colex order.

In the course of the algorithm we define, on the vertex set X, auxiliary hypergraphs $H_1, \ldots, H_n$ and $G_1, \ldots, G_k$ which we need for an estimation of $e(H) = |E(H)|$. The k-hypergraph H_x contains precisely those edges which were added whilst considering vertices from 1 to x inclusive. The i-hypergraph G_i contains as edges those i-subsets B for which the set $\mathcal{T}_B$ was added to H.

We claim that the resulting graph $H = H_n$ is an $\mathcal{F}$-saturated graph. Indeed, H is $\mathcal{F}$-admissible, as we were adding edges only if they did not produce any forbidden subgraphs. On the other hand, take any k-subset E not in $E(H)$. We did not use the opportunity to add E to $E(H)$ when $x = \max E$, $i = k - 1$ and

$A = E \setminus \{x\}$ (when $\mathcal{T}_B = \{E\}$). The only reason for our not doing so is that the addition of E would have created a forbidden subgraph F. Then certainly, $H + E$ contains F, which shows $H \in \mathrm{SAT}(n, \mathcal{F})$.

We claim that $e(G_1) \leq s' - 1$ and

$$e(G_i) \leq (s-1)\binom{n}{i-1}, \quad i = 2, \ldots, k. \tag{13}$$

Assume that for some $i \in [2, k]$ the estimate (13) is not true. Then there is some $(i-1)$-set $V = \{v_1, \ldots, v_{i-1}\}$, $v_1 < \ldots < v_{i-1}$, which is the initial segment of at least s edges of G_i. Let $E_1, \ldots, E_s \in E(G_i)$ be s distinct edges containing V as an initial segment, say $E_j = V \cup \{z_j\}$, $j \in [s]$, $V < z_1 < \ldots < z_s$.

Since $E_1 \in E(G_i)$, all edges whose initial segment is E_1 were added to H at the moment when $x = z_1$ and $A = V$. It follows that $V \notin E(G_{i-1})$ for otherwise these edges would have already been present in H. The only reason that we did not add V to $E(G_{i-1})$ earlier when $x = v_{i-1}$ and $A = \{v_1, \ldots, v_{i-2}\}$ must have been that the hypergraph $H' = H_{v_{i-1}} + \mathcal{T}_V$ contains some forbidden subgraph F. Let

$$Y = \{u \in U_{v_{i-1}} : u \in E \text{ for some } E \in E(F) \cap \mathcal{T}_V\}.$$

As $U_{v_{i-1}}$ is an independent set in H' and each edge in $\mathcal{T}_V$ contains v_{i-1} the assumptions of the theorem imply that $|Y| \leq s$.

By the way algorithm works, any permutation σ of X affecting only the upper shadow U_z of a vertex $z \in X$ (that is, $\sigma(y) = y$ for all $y \leq z$) is an automorphism of H_z because any $\mathcal{T}_B \subset X^{(k)}$ with $z \geq \max B$ is σ-invariant. Applying this remark to $z = v_{i-1}$ we see that we may assume $Y \subset Z = \{z_1, \ldots, z_s\}$.

Let $E \in E(F) \setminus E(H) \subset \mathcal{T}_V$ which exists as $F \not\subset H$. Clearly, $E \cap U_{v_{i-1}} \subset Y$ and $E \in \mathcal{T}_{E_j}$, where $z_j = \min E \cap \{z_1, \ldots, z_s\}$. Since $E_j \in E(G_i)$ we obtain the contradiction $E \in E(H)$, so (13) is proved for any $i \in [2, k]$.

The case $i = 1$ does not fall into general scheme of the proof. But it is rather trivial, for if we have at least s' edges (one-element subsets) in G_1, say $\{v_1\}, \ldots, \{v_{s'}\} \in E(G_1)$, then these vertices span a complete k-graph in H, because if $E \in \{v_1, \ldots, v_{s'}\}^{(k)}$ then $E \in \mathcal{T}_{\{\min E\}} \subset E(H)$. Therefore H contains every k-graph of order s' which is certainly a contradiction.

Clearly, every edge of G_i corresponds to less than $\binom{n-i+1}{k-i}$ edges of H so by (13) we obtain

$$
\begin{aligned}
e(H) - (s' - s)\binom{n}{k-1} &< (s-1)\sum_{i=1}^{k}\binom{n-i+1}{k-i}\binom{n}{i-1} \\
&= 2^{k-1}(s-1)\binom{n}{k-1},
\end{aligned}
$$

which establishes the theorem. ▮

Remark. Our construction is not generally best possible. For example, for $2K_2^2$, the sat-function equals 3 while our algorithm gives $n-1$.

Corollary 2 *For any finite family $\mathcal{F}$ of k-graphs,* $\mathrm{sat}(n, \mathcal{F}) = O(n^{k-1})$. ▮

An interesting question which still remains open is the following.

Problem 3 *Is the estimate* $\mathrm{sat}(n, \mathcal{F}) = O(n^{k-1})$ *valid for any infinite family $\mathcal{F}$ of k-graphs, $k \geq 3$? (True for $k = 2$, see Kászonyi and Tuza [KT86].)*

Tuza [Tuz92] made the following (still open) conjecture which is stronger than (10).

Conjecture 4 (Tuza) *For any k-graph F we have* $\mathrm{sat}(n, F) = \Theta(n^{d(F)})$*, where $d(F)$ is defined by* (6)*. Probably, the stronger assertion* $\mathrm{sat}(n, F) = cn^{d(F)} + O(n^{d(F)-1})$*, for some constant c, is also true.*

4 Variations

Here we consider sat-type questions for a variety of structures. Note that the notion of a saturated structure can be defined in quite general settings, cf. Tuza [Tuz86].

Suppose that we have a class $\mathcal{C}$ of objects with a binary relation '$\subset$' which is a partial order and a rank function $r : \mathcal{C} \to \mathbb{N}$ such that $G \subset H$ implies $r(G) \leq r(H)$. Given a family $\mathcal{F}$ of elements of $\mathcal{C}$, we say that $H \in \mathcal{C}$ is *$\mathcal{F}$-admissible* if H does not contain an $F \in \mathcal{F}$ as a subobject. Now, let $\mathrm{SAT}(n, \mathcal{F})$ be the family of all maximal $\mathcal{F}$-admissible objects of rank n. An object H is called *$\mathcal{F}$-saturated* if $H \in \mathrm{SAT}(r(H), \mathcal{F})$.

In some cases, $\mathcal{C}$ will be the class of hypergraphs with some additional structure: for $G, H \in \mathcal{C}$, $r(H) = v(H)$ and $G \subset H$ holds if G is a subgraph of H in a structure-compatible way. Thus, H is $\mathcal{F}$-saturated if it does not contain any forbidden substructure and this fails to be true for any $H' \in \mathcal{C}$ strictly containing H and having the same order.

Usually, we will ask whether the estimate

$$\mathrm{sat}(n, \mathcal{F}) = O(n^{k-1}) \tag{14}$$

is true for a general 'k-graph' family $\mathcal{F}$ and for the appropriately defined sat-function.

4.1 Graphs with Oriented Edges

Here we shall consider, roughly speaking, k-hypergraphs with the additional structure of directed edges.

4.1.1 Directed Hypergraphs

To obtain a *directed hypergraph* we take a usual hypergraph and on every one of its edges introduce some *orientation*, that is, a linear order.

In fact, estimate (14) is not generally true in these settings. For example, improving previous results of Katona and Szemerédi [KS67], Füredi, Horak, Pareek and Zhu [FHPZ98] showed that $\mathrm{sat}(n, C_3) \approx n \log_2 n$, where $\overrightarrow{C_3}$ denotes the directed 3-cycle: $E(\overrightarrow{C_3}) = \{(1,2), (2,3), (3,1)\}$.

But the situation is different if we consider *cycle-free* (or *acyclic*) hypergraphs, that is, those not containing a *cycle* which is, by definition, an alternating sequence of vertices and edges

$$(x_1, E_1, x_2, E_2, \ldots, x_l, E_l, x_{l+1} = x_1)$$

such that x_i precedes x_{i+1} in E_i. Equivalently, a graph H is cycle-free if we can order its vertices in a way compatible with the ordering of its edges.

By definition, H is *$\mathcal{F}$-saturated* if no $F \in \mathcal{F}$ is a subgraph of H but the addition of any new (ordered) edge to G creates either a forbidden subgraph or an oriented cycle. We say that $A \subset V(F)$ is *independent* if no edge of F lies within A.

Theorem 5 *In the class of the cycle-free k-graphs, let $\mathcal{F}$ be a forbidden family such that the size of any independent set $A \subset F \in \mathcal{F}$ covered by a union of F-edges sharing a vertex outside A, is bounded. Then* $\mathrm{sat}(n, \mathcal{F}) = O(n^{k-1})$.

Proof. We proceed essentially in the same way as in the proof of Theorem 1, but there are new technicalities.

Consider one by one $x \in X = [n]$, $i = 0, \dots, k-1$, $A \in L_x^{(i)}$. Let $B = A \cup \{x\}$ and let $\mathcal{T}_B$ be defined by (12). An orientation of the edges in $\mathcal{T}_B$ is called *symmetric* if any order preserving injections $f, g : [k] \to [n]$ with $f([k]), g([k]) \in \mathcal{T}_B$ induce identical orientations of $[k]$.

If $\mathcal{T}_B \not\subset E(H)$ (as unoriented k-tuples) and there exists a symmetric orientation of $\mathcal{T}_B$ such that $H + \mathcal{T}_B$ does not contain a forbidden subgraph or a cycle, then we add $\mathcal{T}_B$ (with this orientation) to the edge set of H.

That is the algorithm. The obtained hypergraph H does not contain a forbidden configuration. As every k-subset $E \subset X$ was tested (for $B = E$ we had $\mathcal{T}_B = \{E\}$ and every orientation was symmetric), we conclude that $H \in \mathrm{SAT}(n, \mathcal{F})$.

As in Theorem 1 we define the auxiliary hypergraphs H_x (directed) and G_i (undirected). We have to show that $e(G_i) = O(n^{i-1})$.

First, suppose that $E(G_1) = \{\{x_1\}, \dots, \{x_l\}\}$, $x_1 < \dots < x_l$. One can easily check that, as H is cycle-free, there is no choice for the orientation of the edges of $\mathcal{T}_{\{x_i\}}$, $2 \le i \le l$ and H contains the complete cycle-free k-graph on l vertices, which implies $l = O(1)$, as required.

Suppose that $e(G_i) \neq O(n^{i-1})$, for some $1 < i \le k$. Then, for some $(i-1)$-tuple $V \subset X$, we can find an arbitrarily large set $Z = \{z_1, \dots, z_s\} \subset U_x$, $x = \max V$, such that $V \cup \{z_i\} \in E(G_i)$, $i \in [s]$, and the orientation of $\cup_{i \in [s]} \mathcal{F}_{V \cup \{z_i\}} \subset E(H)$ extends to a symmetric orientation '$\prec$' of $\mathcal{T}_V$. As $V \notin E(G_{i-1})$ we conclude that $H' = H_x + (\mathcal{T}_V, \prec)$ contains a forbidden subgraph F or a cycle. If a copy of F is present we follow the proof of Theorem 1. Otherwise let $C = (y_1, E_1, \dots, y_l, E_l, y_{l+1} = y_1)$ be a shortest cycle in H'.

We claim that C can be chosen so that $|W| \le 3k - 5$, where $W = (\cup_{i \in [l]} E_i) \cap U_x$. Then for $s \ge 3k - 5$ we may assume that $W \subset Z$, and the argument of Theorem 1 shows that $C \subset H$, which is a contradiction proving the theorem.

If $Y = \{y_1, \dots, y_l\} \subset U_x$ then $l \le 2$ and the claim is true. Indeed, there is an $i \in [l]$ such that y_{i+1} is larger than y_i and y_{i+2} in $[n]$ but it follows y_i in E_i and precedes y_{i+2} in E_{i+1}, which by the symmetry of $U_x \subset H'$ implies that any two $y, y' \in U_x$ form a 2-cycle.

Next, $|Y \cap U_x| \leq 1$; otherwise pick $y_h, y_i \in U_x \cap Y$, $h < i$, with $y_{i+1} \in Y \setminus U_x$ and obtain a strictly shorter cycle through $(y_1, \ldots, y_h, y_{i+1}, \ldots, y_{l+1} = y_1)$ as $U_x \subset H'$ is 'symmetric'. The two edges containing the point (if it exists) in $Y \cap U_x$ contribute at most $2k-3$ to $|W|$. By the symmetry of U_x, we can assume that for the remaining edges $E_i \cap U_x$ lies within some fixed $(k-2)$-subset of U_x, which shows that $|W| \leq 3k-5$. ∎

For $k = 2$, we can prove a stronger result which includes all infinite families. We exploit the ideas of Kászonyi and Tuza [KT86].

Theorem 6 *In the class of cycle-free* 2*-graphs, we have* $\mathrm{sat}(n, \mathcal{F}) = O(n)$ *for any family* $\mathcal{F}$.

Proof. It is enough to provide a construction. Repeat the following as long as no forbidden subgraph appears: take the next vertex x of $X = [n]$ and add all of $\mathcal{T}_x$. Here, $\mathcal{T}_x$ is the set of the (oriented) edges of the form xy, $y \in U_x$.

Suppose that we have repeated the iteration $m = m(n)$ times. Let $G' = G'(n)$ be the graph received after these m steps. As $[m] \subset V(G')$ spans the complete cycle-free digraph, the number of iterations is bounded by a constant not depending on n; namely, $m < u$, where $u = \min\{v(F) : F \in \mathcal{F}\}$.

Obviously, $m(n)$ is non-increasing as a function of n for $n \geq u$, so it is constant for n sufficiently large. Then, the reason for terminating the procedure is that the addition of $\mathcal{T}_{m+1}$ would create a forbidden subgraph F and it will be the case for any subsequent n, that is, $G'(n) + \mathcal{T}_{m+1}$ contains the same subgraph F.

Now we add edges to G' in any order as long as we create neither a cycle nor a forbidden subgraph. In the resulting graph G, no $d = |V(F) \cap U_{m+1}|$ edges can start at the same vertex $y \in U_m$, as otherwise we have a subgraph isomorphic to F. So, the number of edges in G is at most

$$m(n-1) - \binom{m}{2} + (n-m)(d-1) = O(n). \blacksquare$$

Actually, one can argue that, for sufficiently large n,

$$m = \min\{v(F) - \alpha'(F) : F \in \mathcal{F}\} - 1,$$

where $\alpha'(F)$ is the maximum size of $A \subset V(F)$ such that no edge starts in A. Equivalently, m is the minimum number of vertices one needs to remove from some $F \in \mathcal{F}$ to obtain a *directed star* (a digraph whose edges start at a common vertex). We can take for d the size of any such star. This observation allows us to write more explicitly the bound of Theorem 6.

4.1.2 Ordered Hypergraphs

We can introduce yet another interesting class: *ordered k-graphs*. Every ordered k-graph is a usual (unoriented) k-graph with an extra structure: we have a fixed ordering on the vertex set and the vertices of a subgraph inherit their order from the original graph. To avoid a confusion note that an ordered graph comes equipped with a *fixed* vertex ordering while a cycle-free graph is one that admits at least one compatible vertex ordering.

Without any difficulties we can restate word by word the proof of Theorem 1 (except that now we have already been given an order on the vertex set and in the construction we take the vertices in this order).

Theorem 7 *Let $\mathcal{F}$ be a family of ordered k-graphs. Suppose that there is $s \in \mathbb{N}$ such that the following holds for any $x \in F \in \mathcal{F}$: if $U_x \subset V(F)$ is an independent set covered by a union of F-edges sharing some vertex $y \le x$, then $|U_x| \le s$.*

Then we have $\mathrm{sat}(n, \mathcal{F}) = O(n^{k-1})$. ∎

Using the ideas of Theorem 6 one can see that, for $k = 2$, our result can be extended to all infinite families.

Theorem 8 *For any family $\mathcal{F}$ of ordered 2-graphs,* $\mathrm{sat}(n, \mathcal{F}) = O(n)$. ∎

Trivial examples show that if we enlarge any of the above classes by admitting multiple and/or non-uniform edges, then the estimate (14) fails to be true.

4.2 Layered Hypergraphs

Let $t \in \mathbb{N}$ be fixed. A *layered set* $\mathbf{X}$ of *signature* $\mathbf{n} = (n_1, \ldots, n_t)$ (or an $\mathbf{n}$-*set*) is a sequence of t disjoint sets, $\mathbf{X} = (X_1, \ldots, X_t)$ such that $|X_i| = n_i$, $i \in [t]$. (Usually we typeset symbols in bold when we want to emphasize that the object has some layered structure.) The components of $\mathbf{X}$ are called *layers*.

Given $\mathbf{k} = (k_1, \ldots, k_t)$, a *layered* $\mathbf{k}$*-graph* $\mathbf{G}$ is a pair $(V(\mathbf{G}), E(\mathbf{G}))$, where $V(\mathbf{G})$ is a layered set and $E(\mathbf{G}) \subset (V(\mathbf{G}))^{(\mathbf{k})}$, that is, $E(\mathbf{G})$ is a family of $\mathbf{k}$-subsets of $V(\mathbf{G})$. In other words, every $\mathbf{k}$-graph $\mathbf{G}$ is a k-graph (usually, given $\mathbf{k}$, we denote $k = k_{[t]} = \sum_{i \in [t]} k_i$, etc.) which comes with a fixed partition of the vertex set into t layers such that every edge intersects the ith layer in exactly k_i vertices. The sequence $\mathbf{k}$ is called the *signature* of $\mathbf{G}$; the ith layer of $\mathbf{G}$ is denoted by $V_i(\mathbf{G})$. For example, a bipartite graph is a layered graph of signature $(1,1)$ and, for $t = 1$, we obtain the usual notion of a k-graph. All morphisms between $\mathbf{k}$-graphs preserve layers.

In the obvious way we define the notion of a subgraph, a saturated graph, etc. For example, $\mathrm{SAT}(\mathbf{n}, \mathcal{F})$ consists of all maximal $\mathcal{F}$-admissible $\mathbf{k}$-graphs on a set of signature $\mathbf{n}$.

It is not very hard to extend Theorem 1 to layered graphs. But, to make this work self-contained, we present a complete proof.

For a $\mathbf{k}$-graph $\mathbf{F}$ on $\mathbf{X} = (X_1, \ldots, X_t)$, a set $A \subset X_j$ is called *independent* if for every $\mathbf{E} \in E(\mathbf{F})$, $E_j \not\subset A$.

Theorem 9 *If, for a given family $\mathcal{F}$ of $\mathbf{k}$-graphs, there exists s such that*

1. *for every $\mathbf{F} \in \mathcal{F}$, any independent $A \subset V_1(\mathbf{F})$ covered by a union of $\mathcal{F}$-edges sharing a vertex in $V_1(\mathbf{F}) \setminus A$, has at most s elements;*
2. *for every $j \in [2, t]$ and $\mathbf{F} \in \mathcal{F}$, no $(s+1)$-set $A \subset V_j(\mathbf{F})$ can be covered by a set of $\mathbf{F}$-edges coinciding on the first $j - 1$ layers;*

then there exists $c = c(\mathcal{F})$ such that, for any $\mathbf{n}$,

$$\mathrm{sat}(\mathbf{n}, \mathcal{F}) \leq c \frac{n_1^{k_1} \times \ldots n_t^{k_t}}{\min(n_1, \ldots, n_t)}.$$

Proof. As in Theorem 1, we provide a construction of $\mathbf{H} \in \mathrm{SAT}(\mathbf{n}, \mathcal{F})$.

Order linearly the vertex set $\mathbf{X} = (X_1, \ldots, X_t)$ so that any vertex of X_i comes before any vertex X_j for $i < j$. As usual by $U_x = \{y \in X : y > x\}$ we denote the *upper shadow* of x.

We construct an $\mathcal{F}$-saturated graph $\mathbf{H}$ by starting with the empty $\mathbf{k}$-graph $\mathbf{H}$ on $\mathbf{X}$ and applying the following procedure.

Let j run from 1 to t. Take $x \in X_j$ in order. For every such x let i vary from 0 to $k_j - 1$. Choose one by one $C \subset X_j \setminus U_x$ of size i and let $B = C \cup \{x\}$. Given

B consider in any order sets A such that A intersects every X_l in k_l vertices, $l \in [j-1]$, $A \cap X_j = B$ and $A \cap U_x = \emptyset$. For every such A we consider $\mathcal{T}_A$ which is by the definition the family of $\mathbf{k}$-subsets having A as an initial segment. If $\mathcal{T}_A \not\subset E(\mathbf{H})$ and the addition of the elements of $\mathcal{T}_A$ to the edge set of $\mathbf{H}$ does not create any forbidden subgraph, we add $\mathcal{T}_A$ to $\mathbf{H}$.

We argue that $\mathbf{H}$ exhibits the claimed upper bound in a similar way as in Theorem 1. It is not hard to do, although there are a few new technicalities to overcome.

We define auxiliary $\mathbf{k}$-graphs $\mathbf{H}_1, \ldots, \mathbf{H}_n$ on $\mathbf{X}$ and auxiliary layered graphs $\mathbf{G}_{ji}$ of signature $(k_1, \ldots, k_{j-1}, i)$ on the set $X_1 \cup \ldots \cup X_j$, $j \in [t]$, $i \in [k_j]$.

We need these graphs for estimates of $e(\mathbf{H}) = |E(\mathbf{H})|$. $\mathbf{H}_x$ is the $\mathbf{k}$-graph containing precisely those edges which were added while considering vertices from 1 to x inclusive. The hypergraph $\mathbf{G}_{ji}$ contains as edges those $\mathbf{A} = (A_1, \ldots, A_j)$ for which $|A_j| = i$ and the set $\mathcal{T}_{\mathbf{A}}$ was added to $\mathbf{H}$.

We claim that the resulting graph $\mathbf{H}$ is $\mathcal{F}$-saturated. Indeed, $\mathbf{H}$ is $\mathcal{F}$-admissible, as we were adding edges only if it did not produce any forbidden subgraph. On the other hand, take any edge $\mathbf{E}$ in the complement of $E(\mathbf{H})$. We did not add $\mathbf{E}$ to $E(\mathbf{H})$ when $x = \max \mathbf{E}$, $j = t$, $i = k_t - 1$, $A_l = E_l$ for $l \in [t-1]$ and $A_t = E_t \setminus \{x\}$ (then $\mathcal{T}_{\mathbf{A}} = \{\mathbf{E}\}$). The only reason for this is that it would have created a forbidden subgraph $\mathbf{F}$. Then $\mathbf{H} + \mathbf{E}$ contains $\mathbf{F}$, which shows that $\mathbf{H} \in \mathrm{SAT}(\mathbf{n}, \mathcal{F})$.

We want to show that

$$e(\mathbf{G}_{ji}) \le (s-1)\binom{n_j}{i-1}\prod_{l=1}^{j-1}\binom{n_l}{k_l}, \qquad j \in [t],\ i \in [k_j]. \tag{15}$$

(In fact, $e(G_{11})$ is bounded by some other constant $s' = s'(\mathcal{F})$ but nothing prevents us from assuming $s \ge s'$.) This would establish the theorem as then we would obtain the required

$$\begin{aligned} e(\mathbf{H}) &\le \sum_{j=1}^{t}\sum_{i=1}^{k_j} e(\mathbf{G}_{ji})\binom{n_j}{k_j - i}\prod_{l=j+1}^{t}\binom{n_l}{k_l} \\ &\le \sum_{j=1}^{t}\sum_{i=1}^{k_j}\left((s-1)\binom{n_j}{i-1}\prod_{l=1}^{j-1}\binom{n_l}{k_l}\right)\binom{n_j}{k_j - i}\prod_{l=j+1}^{t}\binom{n_l}{k_l} \\ &= \sum_{j=1}^{t} O\left(\frac{n_1^{k_1}\times\ldots\times n_t^{k_t}}{n_j}\right). \end{aligned}$$

Assume that, for some j and i, estimate (15) is not true. Assume first that $i \neq 1$.

For every edge $\mathbf{E}$ in $\mathbf{G}_{ji}$ consider the set $\mathbf{V}$ of its first $k_1 + \ldots + k_{j-1} + i - 1$ vertices. When $\mathbf{E}$ varies over all edges of $\mathbf{G}_{ji}$, by the pigeon-hole principle some set $\mathbf{V}$ appears at least

$$\left\lceil e(\mathbf{G}_{ji}) \left(\binom{n_j}{i-1} \prod_{l=1}^{j-1} \binom{n_l}{k_l} \right)^{-1} \right\rceil \geq s$$

times. Let $\mathbf{V}$ consist of classes $V_1, \ldots, V_j$ of sizes $k_1, \ldots, k_{j-1}, i-1$ respectively.

Let $\mathbf{E}_1, \ldots, \mathbf{E}_s \in E(\mathbf{G}_{ji})$ be s distinct edges of $\mathbf{G}_{ji}$ containing $\mathbf{V}$ as an initial segment, say $\mathbf{E}_l = \mathbf{V} \cup \{z_l\}$, $l = 1, \ldots, s$, $\mathbf{V} < z_1 < \ldots < z_s$. Let $z = \max \mathbf{V}$.

Since $\mathbf{E}_1 \in E(\mathbf{G}_{ji})$, all edges whose initial segment is $\mathbf{E}_1$ were added to $\mathbf{H}$ at the moment when $x = z_1$, $\mathbf{A} = \mathbf{V} \cup \{z_1\}$. It follows that $\mathbf{V} \notin E(\mathbf{G}_{j,i-1})$, for otherwise these edges would have already been present in $\mathbf{H}$. The only reason that we did not add $\mathbf{V}$ to $E(\mathbf{G}_{ji})$ earlier, when $x = z$, $C = V_j \setminus \{x\}$ and $\mathbf{A} = \mathbf{V}$, must have been that the $\mathbf{k}$-graph $\mathbf{H}' = \mathbf{H}_x + \mathcal{T}_{\mathbf{V}}$ contains some forbidden subgraph $\mathbf{F} \in \mathcal{F}$. Let

$$A = \{u \in X_j \cap U_z : u \in \mathbf{E} \text{ for some } \mathbf{E} \in \mathcal{T}_{\mathbf{V}} \cap E(\mathbf{F})\}. \tag{16}$$

By Assumption 1 (for $j = 1$) or by Assumption 2 (for $j \geq 2$) of the theorem, $|A| \leq s$. One can argue that any layer-preserving permutation σ of $\mathbf{X}$ affecting only U_z is an automorphism of $\mathbf{H}_z$, because any $\mathcal{T}_{\mathbf{B}}$ with $z \geq \max \mathbf{B}$ is σ-invariant. Therefore, we may assume that $A \subset Z = \{z_1, \ldots, z_s\}$.

Now let $\mathbf{E} \in E(\mathbf{F}) \setminus E(\mathbf{H}) \subset \mathcal{T}_{\mathbf{V}}$. Clearly, $\mathbf{E} \in \mathcal{T}_{\mathbf{E}_l} \subset E(\mathbf{H})$, where $z_l = \min(\mathbf{E} \cap \{z_1, \ldots, z_s\})$, since $\mathbf{E}_l \in E(\mathbf{G}_{ji})$; the obtained contradiction $\mathbf{F} \subset \mathbf{H}$ proves (15) for $j \in [t]$, $i \in [2, k_j]$.

Suppose that (15) is not true for $i = 1$. Then as before we argue that there are at least s edges in $\mathbf{G}_{j1}$, say $\mathbf{V}_1, \ldots, \mathbf{V}_s \in E(\mathbf{G}_{j1})$, such that their restrictions to $X_1 \cup \ldots \cup X_{j-1}$ are the same which we denote by $\mathbf{V}$. Let $\mathbf{V}_l \cap X_j = \{v_l\}$, $l \in [s]$.

First, if $j > 1$ then as above we argue that $\mathbf{V}$ is not in $\mathbf{G}_{j-1,k_{j-1}}$ because $\mathcal{T}_{\mathbf{V}_1} \subset \mathcal{T}_{\mathbf{V}}$ was added later. The only reason for omitting $\mathbf{V}$ is that the addition of $\mathcal{T}_{\mathbf{V}}$ would have created a forbidden $\mathbf{F}$. The set A defined by (16) has at most

s elements by Assumption 2; we can assume that $A \subset \{v_1, \ldots, v_s\}$ and deduce a contradiction.

Finally, if $j = 1$ then $\mathbf{H}$ contains all $\mathbf{k}$-edges $\mathbf{E}$ intersecting $\{v_1, \ldots, v_s\}$ and $s = O(1)$ follows. ▮

A version for bipartite graphs (that is, $(1, 1)$-graphs) covers all (including infinite) families and uses slightly different ideas.

Theorem 10 *For any family $\mathcal{F}$ of bipartite graphs, there is $c = c(\mathcal{F})$ such that, for any $n_1, n_2 > 0$,*

$$\mathrm{sat}(n_1, n_2, \mathcal{F}) \leq c \frac{n_1 n_2}{\min(n_1, n_2)}.$$

Proof. Suppose first that $n_1 \geq n_2$. Choose a large $s = s(\mathcal{F})$ (to be specified later). If $n_2 < s$ then any (n_1, n_2)-bipartite graph contains $O(n_1)$ vertices and we are home. Otherwise, as long as no forbidden subgraph appears, take one by one vertices in the first layer and for every such vertex $x \in X_1 = [n_1]$ add all edges connecting it to X_2 to obtain a graph H'. Suppose we do it m times. Note that as $n_2 \to \infty$ then $m = m(n_2)$ does not increase so we can assume that m is constant for every $n_2 \geq s$, some $s = s(\mathcal{F})$. Then, the only thing preventing us from adding the edges $\{\{m + 1, y\} : y \in X_2\}$ is the creation of a forbidden subgraph F. Let $|V(F) \cap X_2| = l$. We see that if we draw through any point $x \in X_1 \setminus [m]$ any l edges, we would obtain a copy of F. Therefore, in whatever way we complete H' to $H \in \mathrm{SAT}(n_1, n_2, \mathcal{F})$, we would have

$$e(H) \leq m n_2 + l n_1 \leq (l + m) n_1 = O(n_1).$$

We settle the case $n_1 \leq n_2$ in the same manner. ▮

4.3 Forbidden Matrices

Here we investigate sat-type problems for 01-matrices. We show that $\mathrm{sat}(n, \mathcal{F}) = O(n^{k-1})$ for any family $\mathcal{F}$ of k-row matrices and indicate other results.

The expression '$n \times m$-*matrix*' means a matrix with n rows (which we view as horizontal arrays) and m vertical columns. We restrict entries to only two values, 0 and 1. For an $n \times m$-matrix M, its *order* $v(M) = n$ is the number of rows and its *size* $e(M) = m$ is the number of columns. Please distinguish expressions like 'an n-row matrix' and 'an n-row' standing respectively for a matrix with n rows and for a row containing n elements.

A matrix F is a *submatrix* of a matrix A (denoted $F \subset A$) if deleting some set of rows and columns of A we can obtain a matrix which is a row/column permutation of F. Given a family $\mathcal{F}$ of matrices (referred to as *forbidden*), we say that a matrix M is *$\mathcal{F}$-admissible* (or *$\mathcal{F}$-free*) if M contains no $F \in \mathcal{F}$ as a submatrix. A *simple matrix* M (that is, a matrix without repeated columns) is called *$\mathcal{F}$-saturated* if M is $\mathcal{F}$-admissible but the addition of any column not present in M violates this property; this is denoted by $M \in \mathrm{SAT}(n, \mathcal{F})$, $n = v(M)$. Please note that, although the definition requires that M is simple, we allow multiple columns in matrices belonging to $\mathcal{F}$.

A popular extremal problem is to consider $\mathrm{forb}(n, \mathcal{F})$, the maximum size of a simple $\mathcal{F}$-admissible matrix with n rows or, equivalently, the maximal size of $M \in \mathrm{SAT}(n, \mathcal{F})$. For example, the fundamental formula (17) falls into this class. The interested reader may start with a recent paper by Anstee, Griggs and Sali [AGS97] containing many references.

On the other hand, the 'dual' of the forb-type problem has received little attention so far. Namely, one can ask what is the value of $\mathrm{sat}(n, \mathcal{F})$, the minimal size of an $\mathcal{F}$-saturated matrix with n rows:

$$\mathrm{sat}(n, \mathcal{F}) = \min\{e(M) : M \in \mathrm{SAT}(n, \mathcal{F})\}.$$

We will be mainly interested in this function. Obviously, $\mathrm{sat}(n, \mathcal{F}) \leq \mathrm{forb}(n, \mathcal{F})$. If $\mathcal{F} = \{F\}$ consists of a single forbidden matrix F then we write $\mathrm{SAT}(n, F) = \mathrm{SAT}(n, \{F\})$, etc.

For an $n \times m$-matrix M and sets $A \subset [n]$ and $B \subset [m]$, $M(A, B)$ denotes the corresponding $|A| \times |B|$-submatrix of M. We use the following self-obvious short-hands: $M(A,) = M(A, [m])$, $M(A, i) = M(A, \{i\})$, etc. For example, the rows and the columns of M are denoted by $M(1,), \ldots, M(n,)$ and $M(,1), \ldots, M(,m)$ respectively while individual entries—by $M(i,j)$, $i \in [n]$, $j \in [m]$.

The $n \times (m_1 + m_2)$-matrix $[M_1, M_2]$ is obtained by concatenating an $n \times m_1$-matrix M_1 and an $n \times m_2$-matrix M_2. Let $mM = [M, \ldots, M]$ denote m copies of M. We write $N \cong M$ to say that N is a column/row permutation of M. Thus, $N \subset M$ if $N \cong M(A, B)$ for some index sets A and B.

By T_k^l we denote the simple $k \times \binom{k}{l}$-matrix consisting of all k-columns with exactly l ones and by K_k—the $k \times 2^k$ matrix of all possible columns of size k. Naturally, $T_k^{\le l}$ denotes the $k \times \binom{k}{\le l}$-matrix consisting of all distinct columns with at most l ones, etc. (We use the shortcut $\binom{k}{\le l} = \binom{k}{0} + \binom{k}{1} + \ldots + \binom{k}{l}$.)

We will need the following result proved independently by Vapnik and Chervonenkis [VC71], Perles and Shelah (see [She72]) and Sauer [Sau73].

$$\mathrm{forb}(n, K_k) = \binom{n}{\le k-1} = \sum_{i=0}^{k-1} \binom{n}{i}. \tag{17}$$

Suppose that $\mathcal{F}$ consists of k-row matrices. Is there any good general upper bound on $\mathrm{forb}(n, \mathcal{F})$ or $\mathrm{sat}(n, \mathcal{F})$? There were different papers dealing with general upper bounds on $\mathrm{forb}(n, \mathcal{F})$, e.g. by Anstee and Füredi [AF86], by Frankl, Füredi and Pach [FFP87] and by Anstee [Ans95], until the conjecture of Anstee and Füredi [AF86] that $\mathrm{forb}(m, \mathcal{F}) = O(n^k)$ for any fixed $\mathcal{F}$ was elegantly proved by Füredi (see [AGS97] for a proof).

On the other hand, we can show that $\mathrm{sat}(n, \mathcal{F}) = O(n^{k-1})$ for any family $\mathcal{F}$ of k-row matrices (including infinite families). Note that we cannot decrease the exponent of $k-1$ with the estimate remaining true for any $\mathcal{F}$; for example, $\mathrm{sat}(n, T_k^k) = \binom{n}{\le k-1}$ as $T_n^{<k}$ is the only matrix in $\mathrm{SAT}(n, T_k^k)$.

Theorem 11 *For any family $\mathcal{F}$ of k-row matrices,* $\mathrm{sat}(n, \mathcal{F}) = O(n^{k-1})$.

Proof. We may assume that K_k is $\mathcal{F}$-admissible for otherwise we are home by (17) as then $\mathrm{sat}(n, \mathcal{F}) \le \mathrm{forb}(n, K_k) = O(n^{k-1})$.

Let $l \in [0, k]$ be the smallest number such that there exists m for which $[mT_k^{\le l}, T_k^{>l}]$ is not $\mathcal{F}$-admissible. Clearly, l is well-defined as, for $l = k$, we obtain the matrix mK_k which, of course, is not $\mathcal{F}$-admissible for large m.

Let $d \ge 1$ be the maximal integer such that $[mT_k^{<l}, dT_k^l, T_k^{>l}]$ is $\mathcal{F}$-admissible for any m. Observe that letting d equal 1 we obtain the matrix $[mT_k^{<l}, T_k^{\ge l}]$ which is $\mathcal{F}$-admissible. Indeed, for $l > 0$ this is true by the choice of l; for $l = 0$ we have K_k which is $\mathcal{F}$-admissible by our assumption. By the choice of l, d is bounded, that is, d is well-defined.

Choose any m such that $[mT_k^{<l}, (d+1)T_k^l, T_k^{>l}]$ is not $\mathcal{F}$-admissible.

Suppose first that $l < k$. Given n, let $N \subset T_n^{l+1}$ be the n-row matrix corresponding to the following set system:

$$H = \bigcup_{j \in [d]} \{Y \in [n]^{(l+1)} : \textstyle\sum_{y \in Y} y \equiv j \pmod{n}\}.$$

Note that any $A \in [n]^{(l)}$ is covered by at most d edges of H as there are at most d possibilities to choose $i \in [n] \setminus A$ so that $A \cup \{i\} \in H$: $i \equiv j - \sum_{a\in A} a \pmod n$, $j \in [d]$.

On the other hand, the set H_1 of all l-subsets of $[n]$ covered by fewer than d edges of H has size at most $2d\binom{n}{l-1}$. Indeed, if $A \in H_1$ then, for some $j \in [d]$ and $x \in A$, $2x = j - \sum_{a\in A-x} a \pmod n$ so, once $A\setminus\{x\}$ and j have been chosen, there are at most 2 choices for x.

Call $X \in [n]^{(k)}$ *bad* if, for some $A \in X^{(l)}$,

$$|\{Y \in H : Y \supset A,\ Y \cap (X \setminus A) = \emptyset\}| \le d - 1. \tag{18}$$

To obtain a bad k-set X, we either complete some $A \in H_1$ to any k-set or take any l-set A and let $X \supset A$ intersect some H-edge covering A. Therefore, the number of bad sets is at most

$$2d\binom{n}{l-1}\binom{n}{k-l} + \binom{n}{l}d\binom{n}{k-l-1} = O(n^{k-1}).$$

Assume that n is so large that $N(X,) \supset mT_k^{<l}$ for any $X \in [n]^{(k)}$. This is possible as $d \ge 1$. Of course, $e(N) = O(n^{k-1})$.

Clearly, $N(X,) \subset [d\binom{n}{l}T_k^{<l}, dT_k^l, T_k^{l+1}]$, for any $X \in [n]^{(k)}$. Hence, N cannot contain a forbidden submatrix by the choice of l and d. Now complete it to an arbitrary $M = [N, N_1] \in \mathrm{SAT}(n, \mathcal{F})$.

Suppose that $e(N_1) \ne O(n^{k-1})$. Then, by (17), $K_k \cong N_1(X,Y)$ for some X, Y. Now, remove the columns corresponding to Y from N_1 and repeat the procedure as long as possible to obtain more than $O(n^{k-1})$ column-disjoint copies of K_k in N_1. If some $X \in [n]^{(k)}$ appears more than d times, then $M(X,) \supset [mT_k^{<l}, (d+1)K_k]$ is not $\mathcal{F}$-admissible. Otherwise, $K_k \subset N_1(X,)$ for some *good* (ie. not bad) $X \in [n]^{(k)}$; but then $N(X,) \supset dT_k^l$ and

$$M(X,) \supset [mT_k^{<l}, dT_k^l, K_k]$$

contains a forbidden matrix. This contradiction proves the required bound for $l < k$.

Let us consider the case when $l = l(\mathcal{F})$ equals k; the above argument does not work in this case because N has size $\Theta(n^k)$, which is too large.

Consider the family $\mathcal{H}$ obtained by interchanging zeros and ones in each $F \in \mathcal{F}$. Clearly, $\mathrm{sat}(n, \mathcal{H}) = \mathrm{sat}(n, \mathcal{F})$. If $l(\mathcal{H}) < k$, then we are home by the above argument applied to $\mathcal{H}$. So, we assume that $l(\mathcal{H}) = k$.

Consider first the case $k = 1$. Let $F \in \mathcal{F}$ be a matrix of the smallest size f. Let the only row of F consist of f_0 zeros and f_1 ones; $f_0 + f_1 = f$. Note that $f_1 \geq 2$ and $f_0 \geq 2$, because $l(\mathcal{F}) = l(\mathcal{H}) = 1$. Trivially, for any n there exists a simple $n \times (f-1)$-matrix M such each row of M contains exactly f_0 zeros. By the minimality of f, M is $\mathcal{F}$-admissible. When we try to complete M to any $\mathcal{F}$-saturated matrix, any added column cannot contain an entry equal to 1; hence, all we can add is at most one all-zero column. Hence, $\mathrm{sat}(n, \mathcal{F}) \leq f$ for any n, which implies the required.

So assume that $k \geq 2$. Now we repeat a part of the above proof with some modifications. Probably, it would be possible to write a general single argument covering all the cases, but we are afraid that the proof would be very hard to follow then.

Let $l' \in [0, k-1]$ be the smallest number such that there exists m for which $[mT_k^{\leq l'}, T_k^{>l'}, T_k^{k-1}, mT_k^k]$ is not $\mathcal{F}$-admissible. Observe that l' is well-defined as this matrix contains mK_k as a submatrix if we let $l' = k-1$.

Define d to be the maximal integer such that $[mT_k^{<l'}, dT_k^{l'}, T_k^{>l'}, T_k^{k-1}, mT_k^k]$ is $\mathcal{F}$-admissible for any m. Note that letting $d = 1$ we obtain the matrix $[mT_k^{<l'}, T_k^{\geq l'}, T_k^{k-1}, mT_k^k]$ which does not contain a forbidden submatrix. Indeed, if $l' > 0$, this is true by the choice of l'; if $l' = 0$, then our matrix $[K_k, T_k^{k-1}, mT_k^k]$ is necessarily $\mathcal{F}$-admissible as $l(\mathcal{H}) = k > 1$ by our assumption.

Choose any m such that $[mT_k^{<l'}, (d+1)T_k^{l'}, T_k^{>l'}, T_k^{k-1}, mT_k^k]$ is not $\mathcal{F}$-free.

Let N be the n-row matrix corresponding to the following set system:

$$H = \bigcup_{j \in [d]} \{Y \in [n]^{(l'+1)} : \textstyle\sum_{y \in Y} y \equiv j \pmod{n}\}.$$

As above we observe that every $A \in [n]^{(l')}$ is covered by at most d edges of H and the number of *bad* sets (that is, such $X \in [n]^{(k)}$ that (18) holds for some $A \in X^{(l')}$) is $O(n^{k-1})$. Assume that n is so large that $N(X,) \supset mT_k^{<l'}$ for any $X \in [n]^{(k)}$, which is possible as $d \geq 1$.

Let $M_1 = [N, T_n^{\geq n-1}]$. Clearly,

$$M_1(X,) \subset [d\tbinom{n}{l'}T_k^{<l'}, dT_k^{l'}, T_k^{l'+1}, T_k^{k-1}, nT_k^k], \quad \text{for any } X \in [n]^{(k)}.$$

Hence, M_1 cannot contain a forbidden submatrix by the choice of l' and d. Now complete it to an arbitrary $M = [M_1, M_2] \in \mathrm{SAT}(n, \mathcal{F})$.

Clearly, $e(M_1) = O(n^{k-1})$. Suppose that $e(M_2) \neq O(n^{k-1})$. Then, by (17), $K_k \cong M_2(X, Y)$ for some X, Y. Now, remove the columns corresponding to Y from M_2 and repeat the procedure as long as possible to obtain more than $O(n^{k-1})$ column-disjoint copies of K_k in M_2. If some $X \in [n]^{(k)}$ appears more than d times then $M(X,) \supset [mT_k^{<l'}, (d+1)K_k, T_k^{k-1}, mT_k^k]$ is not $\mathcal{F}$-admissible. (We assume $n \geq m + k$.) Otherwise, $K_k \subset M_2(X,)$ for some *good* (ie. not bad) $X \in [n]^{(k)}$; but then $N(X,) \supset dT_k^{l'}$ and $M(X,) \supset [mT_k^{<l'}, dT_k^{l'}, T_k^{k-1}, mT_k^k, K_k]$ contains a forbidden matrix. This contradiction proves the theorem. ∎

Let us present some other results.

The following simple observation is useful in tackling sat-type problems. Suppose that no forbidden matrix has two equal rows. Let M' be obtained from $M \in \mathrm{SAT}(n, \mathcal{F})$ by *duplicating* the nth row of M, that is, we let $M'([n],) = M$ and $M'(n+1,) = M(n,)$. Complete M', in an arbitrary way, to an $\mathcal{F}$-saturated matrix. Let C be any added $(n+1)$-column. As both $M'([n],)$ and $M'([n-1] \cup \{n+1\},)$ are equal to $M \in \mathrm{SAT}(n, \mathcal{F})$, we conclude that both $C([n])$ and $C([n-1] \cup \{n+1\})$ must be columns of M. As C is not an M'-column, $C = (C', b, 1-b)$ for some $(n-1)$-column C' such that both $(C', 0)$ and $(C', 1)$ are columns of M. This implies that $\mathrm{sat}(n+1, \mathcal{F}) \leq e(M) + 2l$, where l is the number of pairs of equal columns in M after we delete the nth row. In particular, the following theorem follows.

Theorem 12 *Suppose that no matrix in $\mathcal{F}$ has two equal rows. Then either $\mathrm{sat}(n, \mathcal{F})$ is constant for large n or $\mathrm{sat}(n, \mathcal{F}) \geq n + 1$ for every n.*

Proof. If we have some $M \in \mathrm{SAT}(n, \mathcal{F})$ with at most n columns then a well-known theorem of Bondy [Bon72a] (see e.g. [Bol86, Theorem 2.1]) implies that there is $i \in [n]$ such that the removal of the ith row does not produce multiple columns. Now the duplication of the ith row gives an $\mathcal{F}$-saturated matrix, which implies $\mathrm{sat}(n+1, \mathcal{F}) \leq \mathrm{sat}(n, \mathcal{F})$, and the theorem follows. ∎

There are many open problems concerning particular forbidden matrices; for example, the computation of $\mathrm{sat}(n, T_m^k)$ or $\mathrm{sat}(n, K_k)$. Of course, Theorem 12 is applicable here. While it is easy to see that $\mathrm{sat}(n, T_m^k) \geq n+1$ for any $m \in [0, k]$ and $k \geq 2$, we do not know for which k we have $\mathrm{sat}(n, K_k) = O(1)$. We could only show that $\mathrm{sat}(n, K_2) = n+1$, which is an easy (and perhaps known) result, and (surprisingly) $\mathrm{sat}(n, K_3) = 10$ for $n \geq 4$. We do not provide any proofs here, except we exhibit an example of an n-row K_3-saturated matrix of size 10 for any $n \geq 6$. For $n = 6$ we can take

$$M = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}.$$

It is possible (but rather boring) to check by hand that M is indeed K_3-saturated as is, in fact, any $n \times 10$-matrix M' obtained from M by duplicating any row, cf. Theorem 12. (The symmetries of M shorten the verification.)

The author would like to thank Richard Anstee for drawing the author's attention to the Turán-type problem for matrices and Andrew Thomason for his computer programme used for computing $\mathrm{sat}(n, \mathcal{F})$ for small n and $\mathcal{F}$.

4.4 Edge Killers

Here we introduce certain extremal problems which are closely related to the sat-type questions. We settle the problem for complete 2-graphs, which extends a theorem of Erdős, Hajnal and Moon [EHM64] who computed $\mathrm{sat}(n, K_m^2)$.

Given a forbidden family $\mathcal{F}$, we say that a k-graph G $\mathcal{F}$-*kills* (or simply *kills* when $\mathcal{F}$ is understood) an edge $E \in E(\overline{G})$ if the addition of E to G creates a new forbidden subgraph. For example, $G \in \mathrm{m\text{-}SAT}(n, \mathcal{F})$ if and only if it kills all edges in its complement. The $\mathcal{F}$-*closure* $\mathrm{Cl}^*_{\mathcal{F}}(G)$ of G is the k-graph on $V(G)$ consisting of all edges of G plus all $\mathcal{F}$-killed edges. Let $\mathrm{cl}^*_{\mathcal{F}}(G) = |\mathrm{Cl}^*_{\mathcal{F}}(G)|$.

Let us define $\text{k-m-sat}(e, \mathcal{F}; n)$ to be the maximum size of $\mathrm{Cl}^*_{\mathcal{F}}(G)$ where G is a k-graph of order n and size e, $e \leq \mathrm{m\text{-}sat}(n, F)$. In the same way we define $\text{k-sat}(e, \mathcal{F}; n)$ except we consider only $\mathcal{F}$-free graphs of order n and size e, $e \leq \mathrm{sat}(n, \mathcal{F})$. We agree that k-sat = k-m-sat = $\binom{n}{k}$ for other (larger) values of e. Clearly, $\text{k-m-sat}(e, \mathcal{F}; n) \geq \text{k-sat}(e, \mathcal{F}; n)$; both k-m-sat and k-sat are monotone increasing in e.

Here we compute k-m-sat and k-sat (and describe all extremal graphs) for complete 2-graphs. This extends a result of Erdős, Hajnal and Moon [EHM64] who computed $\text{sat}(n, K_m^2)$, as $\text{sat}(n, \mathcal{F}) = \min\{e : \text{k-sat}(e, \mathcal{F}; n) = \binom{n}{k}\}$.

Let us provide a construction. Given $n \geq m \geq 3$ and e,

$$\binom{m}{2} - 1 \leq e \leq \binom{n}{2} - \binom{n-m+2}{2} = \text{m-sat}(n, K_m^2) = \text{sat}(n, K_m^2),$$

write $e - \binom{m-2}{2} = l(m-2) + r$ with $r \in [0, m-3]$. Choose an $(m-2)$-set A and a disjoint l-set B. Let G be $P_{A,B}$ (which consists of all edges lying within $A \cup B$ and intersecting A) plus any r extra edges, none within B. (So B is an independent set in G.) It is routine to check that G can be accommodated within $[n]$. Clearly, G kills all $\binom{l}{2}$ edges of $K^2(B)$. We show that this is best possible by applying the contraction technique of Erdős, Hajnal and Moon [EHM64].

Theorem 13 *In the above notation,*

$$\text{k-sat}(e, K_m^2; n) = \text{k-m-sat}(e, K_m^2; n) = \binom{l}{2} + e, \tag{19}$$

and all extremal graphs are given by the construction preceding the theorem.

Proof. To prove the upper bound, we use induction on l with the case $l = 2$ being trivially true. Let $l \geq 3$. Given a graph G of order n and size e (not necessarily K_m^2-free), fix any killed edge $\{x_1, x_2\}$ and let G' be obtained from G by contracting the vertices x_1 and x_2 into one vertex x. Fix an $(m-2)$-set Y such that $G[Y \cup \{x_1, x_2\}]$ is the complete graph but for $\{x_1, x_2\}$; colour these $\binom{m}{2} - 1$ edges red. Clearly, during the contraction at least $m-2$ red edges disappear, so $e(G') \leq e(G) - m + 2$.

Obviously, an edge killed by G is also killed by G' (except $\{x_1, x_2\}$) but two G-killed edges, say $\{a, x_1\}, \{a, x_2\} \in E(\overline{G})$, may produce only one edge in G' (which is also killed). When this happens then, for $i = 1, 2$, choose an arbitrary $(m-2)$-set X_i with $G[X_i \cup \{a, x_i\}] = P_{m-2,2}$ and colour all edges connecting a to X_i blue. Let D be a blue edge. Some $a \in D$ is incident neither to x_1 nor to x_2, so D is not coloured red. As the other endvertex of D sends at least one edge to $\{x_1, x_2\}$, D cannot be coloured blue more than twice.

We have $e - \binom{m}{2} + 1$ non-red edges each being coloured blue at most twice, while each time two killed edges contract together exactly $2(m-2)$ edges are coloured blue. This yields

$$\text{cl}^*_{K_m^2}(G) - e(G) \leq \text{cl}^*_{K_m^2}(G') - e(G') + \left\lfloor \frac{e - \binom{m}{2} + 1}{m-2} \right\rfloor + 1. \tag{20}$$

(the last term 1 counts the edge $\{x_1, x_2\}$) and the induction assumption applied to the graph G' of order $n-1$ and size at most $e-m+2$ proves the desired upper bound.

Let us follow our argument to characterize the cases of equality. Clearly, for $l=2$, when $e=\binom{m}{2}-1+r$, we must have an induced $P_{m-2,2}$-subgraph present while the remaining r edges can be placed arbitrarily, which is precisely what our construction says.

Let $l \geq 3$ and let G be an extremal graph. Apply the above contraction to G, preserving the above notation. By induction, $G' = P_{A,B} + E_1 + \ldots + E_r$, where B is an independent $(l-1)$-set disjoint from an $(m-2)$-set A. The vertex x, which has degree at least $m-2$ in G', must belong to $A \cup B$ as $r \leq m-3$.

Suppose that $x \in A$. Then the $(m-1)$-set $Y \cup \{x\}$, which spans the complete graph in G', must equal $A \cup \{y\}$, for some $y \in B$. Each blue edge of G lies within a $P_{m-2,2}$-subgraph in G' and, as $r \leq m-3$, none of $E_1, \ldots, E_r$ can be blue (nor red, of course). But then, for $z \in B \setminus \{y\}$, $E(G)$ contains either $\{x_1, z\}$ or $\{x_2, z\}$ (because $\{x, z\} \in E(G')$) which is also uncoloured. So, we have at least $r+1$ uncoloured edges and we cannot have equality in (20), which is a contradiction.

Hence, $x \in B$; then Y must equal A, and G is given by our construction. ∎

5 Irregularities

Here we demonstrate many irregularities of the sat-function in the comparison to the *Turán function* $\mathrm{ex}(n, \mathcal{F}) = \max\{e(G) : G \in \mathrm{SAT}(n, \mathcal{F})\}$.

Clearly, $\mathrm{ex}(n, F_1) \leq \mathrm{ex}(n, F_2)$ whenever F_1 is a subgraph of F_2. Kászonyi and Tuza [KT86] demonstrated an example of $F_1 \subset F_2$ with $\mathrm{sat}(n, F_1) > \mathrm{sat}(n, F_2)$ for all large n. Tuza [Tuz92, p. 401] asks if there exists a *connected irregular* pair $F_1 \subset F_2$; this is answered in the affirmative by the following simple example.

Example 14 *There is a pair of connected graphs $F_1 \subset F_2$ on the same vertex set such that* $\mathrm{sat}(n, F_1) > \mathrm{sat}(n, F_2)$ *for all* $n \geq v(F_1)$.

Proof. Let $m \geq 5$ and $F_1 = S_m^2$, that is, $V(F_1) = [m]$ and $E(F_1) = \{\{1, i\} : i \in [2, m]\}$ and let F_2 be obtained from F_1 by adding the edge $\{2, 3\}$. Clearly, $\mathrm{sat}(n, F_2) \leq n-1$, $n \geq m$, as S_n^2 is an example of an F_2-saturated graph.

On the other hand, in any monotonically F_1-saturated graph G, any two vertices of degree at most $m-3$ must be connected. (Otherwise the addition of this edge cannot create a forbidden subgraph.) If we have $v \in [0, m-2]$ such vertices, then $e(G) \geq \binom{v}{2} + (m-2)(n-v)/2$, which is easily seen to exceed $n-1$ for all $n \geq m$. ∎

Remark. Curiously enough, the w-sat-function (studied later) exhibits the analogous irregularity on the very same pair: it is not hard to check that $\text{w-sat}(n, F_2) = e(F_2) - 1 = m-2$ while $\text{w-sat}(n, F_1) = \binom{m-1}{2}$, $n \geq m$.

Clearly, for every $n \geq v(F)$, we have $\text{ex}(n, F) \leq \text{ex}(n+1, F)$. On the other hand, Kászonyi and Tuza [KT86] observe that, for any $n = 2k-1$, $\text{sat}(n, P_3) = k+1 > \text{sat}(n+1, P_3) = k$, where P_3 is the path with 3 edges. Our next example amplifies this irregularity.

Example 15 *For every constant d, there is a 2-graph $F = F(d)$ such that*

$$\text{sat}(n, F) < \text{sat}(n \pm 1, F) - d,$$

for a periodic series of values of n.

Proof. Let $m = 2d+3$ and let $F = B_{mm}$ be the *dumb-bell*

$$E(B_{mm}) = [m]^{(2)} \cup [m+1, 2m]^{(2)} \cup \{\{1, m+1\}\},$$

that is, B_{mm} is the disjoint union of two copies of K_m plus one edge connecting them.

Let us show that the claim is true for any $n = lm$ if $l \in \mathbb{N}$ is large. Clearly, $\text{sat}(lm, F) \leq lm(m-1)/2$ (in fact, this is sharp) as $lK_m^2 \in \text{SAT}(lm, F)$. On the other hand, let $n = lm-1$ and suppose that $G \in \text{sat}(n, F)$ has at most $g = lm(m-1)/2 + d$ edges.

Clearly, $\delta(G)$, the minimal degree of G, is at least $\delta(B_{mm}) - 1 = m-2$. Suppose that for some $x \in V(G)$ $d(x) = m-2$. Then for every y non-incident to x the edge $E = \{x, y\} \in E(\overline{G})$ cannot be the bridge in a created B_{mm}-subgraph as the degree of x is too small; that is, x and y fall in the same K_m^2-half. Therefore, y must be connected to all $m-2$ neighbours of x and $e(G) \geq (m-2)n + O(1)$ which is a contradiction.

Hence $\delta(G) \geq m-1$. The inequality $\Delta(G)+(m-1)(n-1) \leq 2e(G) \leq 2g$ implies that $\Delta(G) \leq 2(d+m-1)$. If some $x \in V(G)$ does not belong to an m-clique then any missing edge $\{x, y\}$ must create a K_m^2-subgraph and we arrive at a contradiction again, as $d(x) \leq \Delta(G)$ is bounded. Thus the whole of $V(G)$ is covered by m-cliques.

We want to find a set $X \subset V(G)$ with the *surplus* $s(X) = e(G[X]) - \frac{m-1}{2}|X|$ at least $m-1$ as then the claim would follow:

$$e(G) \geq e(G[X]) + \frac{m-1}{2}(n-|X|) \geq \frac{m-1}{2}n + m - 1 > g.$$

As m does not divide n, there are two distinct cliques $A, B \in V(G)^{(m)}$ with $i = |A \cap B| > 0$. It is straightforward to verify that

$$s(A \cup B) = 2\binom{m}{2} - \binom{i}{2} - \frac{m-1}{2}(2m-i) \geq \frac{m-1}{2}.$$

No m-clique $C \not\subset A \cup B$ can intersect some other clique or $A \cup B$. (Otherwise we gain another suplus of $(m-1)/2$.) By the divisibility argument, $i = 1$. As a $(2m-1)$-clique has suplus at least $m-1$, there exists some $E \in E(\overline{G})$ lying within $A \cup B$. It is easy to see that $G+E$ must contain a K_m^2-subgraph on some m-set $C \not\subset A \cup B$ intersecting $A \cup B$, which implies $s(A \cup B \cup C) \geq m-1$ as required.

Clearly, for $n = ml+1$, $\mathrm{sat}(n, B_{mm}) \geq \frac{m-1}{2}n > g$, which completes the proof. ∎

The elegant averaging argument of Katona, Nemetz and Simonovits [KNS64] shows that the limit $\mathrm{ex}(n, \mathcal{F})/n^k$ exists for any family $\mathcal{F}$ of k-graphs. Concerning the sat-function, Tuza [Tuz88] made the following (still open) conjecture.

Conjecture 16 (Tuza) *For any 2-graph F, the limit $\lim_{n\to\infty} \mathrm{sat}(n, F)/n$ exists.*

We can show that this assertion is not true for families of forbidden graphs.

Example 17 *There exists a finite family $\mathcal{F}$ of 2-graphs such that, for some $c > 0$ and for infinitely many n, $\mathrm{sat}(n, \mathcal{F}) < \mathrm{sat}(n \pm 1, \mathcal{F}) - cn$. In particular, the ratio $\mathrm{sat}(n, \mathcal{F})/n$ does not necessarily tend to a limit for a finite family $\mathcal{F}$ of 2-graphs.*

Proof. Fix $m \geq 4$ and consider the family $\mathcal{F}$ consisting of the dumb-bell B_{mm} and $F_{m1}, \ldots, F_{m,m-1}$, where

$$E(F_{mi}) = [m]^{(2)} \cup [m-i+1, 2m-i]^{(2)}, \quad i \in [m-1],$$

that is, F_{mi} is the union of two K_m^2-graphs sharing i common vertices.

Clearly, the disjoint union of K_m^2-graphs is $\mathcal{F}$-saturated as any missing edge connects two different copies and thus creates a B_{mm}-subgraph. Hence, if m divides n then $\mathrm{sat}(n, \mathcal{F}) \leq \frac{n}{m}\binom{m}{2}$.

On the other hand, suppose that m does not divide n and let G be any $\mathcal{F}$-saturated graph on $[n]$. By the definition of $\mathcal{F}$, no vertex can belong to two different K_m^2-subgraphs of G; suppose that the sets $A_i = [m(i-1)+1, mi]$, $i \in [s]$, are all m-sets spanning complete subgraphs in G.

Note the following two properties of G. *Property A*: $G[A_{[s]}] \cong sK_m^2$. (Because B_{mm} is forbidden.) *Property B*: any missing edge E intersecting $B = [n] \setminus A_{[s]}$ creates a K_m^2-subgraph. (Because it is impossible that $B_{mm} \subset G + E$ with E being the bridge.)

We claim that these two properties and the fact that $B \neq \emptyset$ (as m is not a divisor of n) imply that

$$e(G) \geq \frac{n}{m}\left(\binom{m}{2} + m - 2\right) - m^2. \tag{21}$$

We use induction on n. If some $E \in B^{(2)}$ is not a G-edge then it is easy to check that the graph G' obtained from G by contracting the edge E has the properties in question. The endvertices of E have at least $m-2$ common neighbours in G (because E creates a K_m^2-subgraph) so $e(G) \geq e(G') + m - 2$ and (21) follows by induction.

Suppose that B spans the complete graph in G. If some $E \in E(\overline{G})$ intersects both A_i and B then a K_m^2-subgraph created by E lies within $A_i \cup B$ and so at least $m-2$ G-edges intersect both A_i and B. Therefore,

$$e(G) \geq f(b) = (n-b)\frac{m-1}{2} + \binom{b}{2} + \frac{n-b}{m}(m-2),$$

where $b = |B|$. (We correspondingly count the edges within $A_{[s]}$, within B and in between.) The minimum of f is achieved for $b = \frac{m}{2} + \frac{m-2}{m}$ and our estimate (21) follows rather crudely.

Hence, if we increase/decrease $n = ml$ by one, then $\mathrm{sat}(n, \mathcal{F})$ increases at least by $n\frac{m-2}{m} + O(1)$. ∎

6 Specific Classes

Our aim in this section is to give precise information about $\mathrm{sat}(n, \mathcal{F})$ for special classes $\mathcal{F}$.

6.1 Stars

The *star* $S_m^k = P(1, m-1; 1, k-1)$, $m > k \geq 2$, has $[m]$ as the vertex set and $\{E \in [m]^{(k)} : E \ni m\}$ as the edge set. In other words, S_m^k has m vertices and its $\binom{m-1}{k-1}$ edges are the k-tuples containing some fixed vertex, which is called the *centre*.

The exact values of $\mathrm{sat}(n, S_m^k)$ are known only for S_m^2, any m, (see [KT86]) and for S_4^3 (see [EFT91]).

The asymptotic behaviour of $\mathrm{sat}(n, S_{k+1}^k)$ was found by Erdős, Füredi and Tuza [EFT91, Theorem 2]. Exploiting their ideas we extend their result to all stars; this theorem appears in [Pik99b].

Theorem 18 *Let $m > k \geq 2$ and $S = S_m^k$. Then*

$$\frac{m-k}{2}\binom{n}{k-1} \geq \mathrm{sat}(n, S) \geq \text{m-sat}(n, S) \geq \frac{m-k}{2}\binom{n}{k-1} - O(n^{k-4/3}). \quad (22)$$

Proof. Let us provide a construction of an S-saturated graph $G = G_{m,n}^k$ of order n proving the upper bound. Partition the vertex set $[n]$ into $n' = \lceil n/(m-k+1) \rceil$ *blocks* $B_1, \ldots, B_{n'}$ of size $m-k+1$ each except possibly the last one. The edge set is

$$E(G) = \left\{ F \in [n]^{(k)} : |F \cap B_j| \geq 2,\ j = \min\{i \in [n'] : F \cap B_i \neq \emptyset\} \right\}.$$

Thus every edge of G has at least two common points with some B_j and intersects no B_i with $i < j$.

Let us show that $S \not\subset G$. Suppose not and we have an S-subgraph $S' \subset G$ centered at x. Let

$$j = \min\{i \in [n'] : V(S') \cap B_i \neq \emptyset\}. \quad (23)$$

Choose a k-set $F \ni x$ so that it contains one vertex from B_j and some $k-1$ vertices in $V(S') \setminus B_j$ which is possible since $|V(S') \setminus B_j| \geq k-1$. We obtain a contradiction as on one hand F contains the centre x and must belong to S while on the other hand $F \notin E(G)$ by definition.

If we add any extra edge F to G then the set $Y = F \cup B_j$ spans a copy of S centered at x where B_j is the first block intersecting F and $\{x\} = F \cap B_j$. Indeed, every $F' \in Y^{(k)}$ containing x either equals F or intersects B_j in at least two points and so belongs to $E(G)$.

Therefore we conclude that G is S-saturated. To prove the desired upper bound $|G^k_{m,n}| \leq \frac{m-k}{2}\binom{n}{k-1}$ we observe, for $k=2$, that each vertex of the 2-graph $G^2_{m,n}$ has degree at most $m-k$ while, for $k \geq 3$, we use induction and the equality $|G^k_{m,n+1}| = |G^k_{m,n}| + |G^{k-1}_{m-1,n}|$.

Of course, $\mathrm{sat}(n,S) \geq \text{m-sat}(n,S)$.

Finally, let G be a minimum monotonically S-saturated graph on $V=[n]$. By the definition, the addition to G of any edge $F \in E(\overline{G})$ creates at least one S-subgraph $S' \subset G+F$. Let $\mathcal{S}(F)$ be the set of all such subgraphs S' created by F.

Let $\mathcal{F}(F)$ denote the set of edges in $\overline{G}$ which intersect $F \in E(G)$ in $k-1$ points and create a copy of S containing F as an edge. Formally,

$$\mathcal{F}(F) = \left\{F' \in E(\overline{G}) : |F \cap F'| = k-1,\ \exists S' \in \mathcal{S}(F')\ F \in E(S')\right\}, \quad F \in E(G).$$

Also we define

$$\begin{array}{rcll} \mathcal{F}(G') & = & \bigcup_{F \in E(G')} \mathcal{F}(F), & G' \subset G, \\ \partial F & = & F^{(k-1)}, & F \in [n]^{(k)}, \\ \partial G' & = & \bigcup_{F \in E(G')} \partial F, & \text{a } k\text{-graph } G'. \end{array}$$

As G is monotonically S-saturated we conclude that

$$\mathcal{F}(G) = V^{(k)} \setminus E(G). \tag{24}$$

Choose an integer $t = t(n)$, to be specified later, such that $t \to \infty$ and $t/n \to 0$. On the vertex set V we define two subgraphs $G_0, G_1 \subset G$; G_0 is a maximal subgraph of G with $|\mathcal{F}(G_0)| \leq t|G_0|$ and G_1 consists of the edges of G not in G_0: $E(G_1) = E(G) \setminus E(G_0)$. By the maximality of G_0 for every $F \in E(G_1)$ we have

$$|\mathcal{F}(F) \setminus \mathcal{F}(G_0)| > t. \tag{25}$$

From (24) and the proved upper bound in (22) we conclude that $|\mathcal{F}(G)| = \binom{n}{k} - |G| = \binom{n}{k} - O(n^{k-1})$. Taking into the account that $\mathcal{F}(G) = \mathcal{F}(G_0) \cup \mathcal{F}(G_1)$ and $|\mathcal{F}(G_0)| \le t|G_0| = O(tn^{k-1})$ we obtain

$$|X| = \binom{n}{k} - O(tn^{k-1}), \tag{26}$$

where $X = \mathcal{F}(G_1) \setminus \mathcal{F}(G_0)$.

Let $Z = V^{(k-1)} \setminus \partial G_1$. We claim that

$$|Z| = O(t^{1/2} n^{k-3/2}). \tag{27}$$

Suppose not. Then the average value of $z(D) = |\{E \in Z : E \supset D\}|$ over all $D \in V^{(k-2)}$ is greater than $O(t^{1/2}n^{1/2})$. For any $E, E' \in Z$ with $|E \cap E'| = k-2$ we have $F = E \cup E' \not\in X$, because otherwise at least one of $E, E' \in \partial F$ is covered by an edge of $S' \in \mathcal{S}(F)$ which then is necessarily an edge of G_1 (as it intersects $F \in \mathcal{F}(G_1) \setminus \mathcal{F}(G_0)$ in $k-1$ vertices). Therefore, we have at least $\binom{k}{2}^{-1} \sum_{D \in V^{(k-2)}} \binom{z(D)}{2}$ k-sets not in X, which exceeds $\binom{n}{k-2} O(tn)$ by the convexity of $\binom{x}{2}$. This contradicts (26) and proves the claim.

Let

$$g_1(E) = |\{F \in E(G_1) : F \supset E\}|, \quad E \in \partial G_1.$$

Take any $F \in E(G_1)$. Let $\partial F = \{E_1, \ldots, E_k\}$. We claim that all but at most two of $g_1(E_i)$'s are larger than $t/6$. Suppose not, say $g_1(E_i) \le t/6$, $i = 1, 2, 3$. Take $F' \in \mathcal{F}(F) \setminus \mathcal{F}(G_0)$ and any $S' \in \mathcal{S}(F')$ containing F as an edge. Let $F' = E_i \cup \{x\}$, some $i \in [k]$, $x \in V \setminus F$. The star S' contains $k-2$ edges of the form $E_j \cup \{x\}$, $j \neq i$. These edges cannot be in G_0 and so contribute at least 1 to $g_1(E_1) + g_1(E_2) + g_1(E_3)$. In total, each $\{x\} \cup E_j \in E(G_1)$ is counted at most twice. (Once it occurs then at most 2 edges of the form $\{x\} \cup E_i$ can belong to $E(\overline{G})$.) But this contradicts (25). The claim is proved.

Define

$$\begin{aligned} W &= \{E \in \partial G_1 : g_1(E) \le m - k - 1\}, \\ T &= \{F \in E(G_1) : W \cap \partial F \neq \emptyset\}. \end{aligned}$$

We claim that $|W| = O(t^{1/2} n^{k-3/2})$. Suppose not. Note that for $E, E' \in W$ with $|E \cap E'| = k-2$ we necessarily have $F = E \cup E' \not\in X$ for otherwise in an $S' \in \mathcal{S}(F)$ centered at x, say $x \in E$, there are $m-k$ edges (necessarily in $E(G_1)$)

different from F and covering E. Thus there are at least $\binom{k}{2}^{-1}\sum_{D\in V^{(k-2)}}\binom{w(D)}{2}$ edges not in X, where $w(D) = |\{E \in W : E \supset D\}|$, $D \in V^{(k-2)}$. Using the convexity of the $\binom{x}{2}$-function as before we can argue that there more than $O(tn^{k-1})$ edges not in X, contradicting (26). The claim is established.

Every $E \in W$ is contained in at most $m-k-1$ edges of G_1, so $|T| = O(t^{1/2}n^{k-3/2})$. For every $F \in E(G_1)\setminus T$ we have $\sum_{E\in\partial F}\frac{1}{g_1(E)} \le \frac{2}{m-k} + (k-2)\frac{6}{t}$. Note the following easy identity

$$\begin{aligned}|\partial G_1| &= \sum_{F\in E(G_1)\setminus T}\left(\sum_{E\in\partial F}\frac{1}{g_1(E)}\right) + \sum_{F\in T}\left(\sum_{E\in\partial F}\frac{1}{g_1(E)}\right)\\ &\le \left(\frac{2}{m-k} + O(1/t)\right)|G_1| + k|T|.\end{aligned}$$

We know, see (27), that $|\partial G_1| = \binom{n}{k-1} - O(t^{1/2}n^{k-3/2})$. Hence

$$\frac{m-k}{2}\binom{n}{k-1} - |G| = O(t^{1/2}n^{k-3/2} + |G|/t) = O(t^{1/2}n^{k-3/2} + n^{k-1}/t).$$

Taking $t = \lfloor n^{1/3} \rfloor$ we obtain the required result. ∎

6.2 Sub-Designs

A *t-(v,k,λ)-design* (or an $S_\lambda(t,k,v)$) is a k-graph G of order v in which every t-set is covered by exactly λ edges. As the question whether a design exists for a given set of parameters is generally notoriously hard, one direction of research is to consider what we call here *sub-designs*. A *t-(v,k,λ)-sub-design* G is a maximal k-graph of order v such that no t-set is covered by more than λ edges. Clearly, in the latter case, G can contain at most $\lambda\binom{v}{t}\binom{k}{t}^{-1}$ edges so we take $e(G)$ as the measure of the 'goodness' of G.

It is easy to construct sub-designs. This can be done, for example, by starting with the empty graph and consecutively adding missing k-subsets as long as possible. If we are lucky, we obtain an $S_\lambda(t,k,v)$; in this case $e(G)$ is maximal possible. On the other hand, one can ask how unlucky we can be, that is, how small G can be. Let $\mathcal{D} = \mathcal{D}(\lambda,t,k)$ be the family of all k-graphs with $\lambda+1$ edges such that some t vertices belong to every edge. Then $\mathrm{SAT}(n,\mathcal{D})$ is the family of all sub-designs of order n. Thus we are interested in $\mathrm{sat}(n,\mathcal{D}(\lambda,t,k))$, the minimal size of a t-(n,k,λ)-sub-design.

Note that $\mathcal{D}(\lambda, 1, 2)$ consists of one graph, namely the star $S^2_{\lambda+2} = K_{1,\lambda+1}$. Kászonyi and Tuza [KT86] computed $\mathrm{sat}(n, S^2_{\lambda+2})$. In fact, their method extends to any $\mathcal{D}(\lambda, 1, k)$.

We need the following simple lemma, whose proof we include for the sake of completeness.

Lemma 19 *Given integers n', λ and k with $\lambda \le \binom{n'-1}{k-1}$, there is a k-graph G' on $[n']$ such that every vertex has degree λ except a set D of at most $k-1$ vertices of degree $\lambda - 1$.*

Proof. Place the elements of $[n']$ clockwise on a circle to form a regular n'-gon. Define the equivalence relation $\sim$ on $[n']^{(k)}$ so that two k-sets are equivalent if some rotation maps one onto the other. (Note that we do not allow mirror reflections.) Let $H_1, \ldots, H_p \subset [n']^{(k)}$ be the obtained equivalence classes. Clearly, each H_i is a regular covering of $[n']$ of degree which is a divisor of k. Let H_p be the equivalence class of the set $[k]$ which consists of k consecutive vertices. Starting with the empty hypergraph G' on $[n']$, for $i \in [p-1]$, add H_i to G' if the maximal degree does not exceed λ. At the end, we will be left with some d-regular k-graph. Clearly, $\lambda - d$ is at most k because otherwise we had to add every H_i, $i \in [p-1]$, so adding H_p we obtain the complete k-graph on $[n']$, which implies the contradiction $\binom{n'-1}{k-1} < \lambda$.

Finally, we try to add some subset of H_p to make G' nearly λ-regular. Take some edge $E \in H_p$ which has not been added to G', say $E = [i+1, i+k]$. We add, one by one, the following shifts of E:

$$[i+1, i+k],\ [i+k+1, i+2k],\ [i+2k+1, i+3k],\ \ldots$$

and so on in this order until either we come across E again or we cannot add the current edge (because then the maximal degree of G becomes larger than λ). In the former case, we take any other unused edge and repeat the procedure. In the latter case, we have the required graph built because every time the added portion of H_p is nearly regular, that is, the difference between the maximal and minimal degrees is always at most 1. ∎

The following theorem gives the exact answer in almost every case, except for some small n when we have only a lower bound.

Theorem 20 *Given $\lambda \geq 1$ and $k \geq 2$, let $\mathcal{D} = \mathcal{D}(\lambda, 1, k)$ and define v by $\lambda/k \in [\binom{v-1}{k-1}, \binom{v}{k-1})]$. Then* $\text{m-sat}(n, \mathcal{D}) \geq \binom{v}{k} + \lceil \frac{\lambda(n-v)}{k} \rceil$. *If, furthermore, $\lambda \leq \binom{n-v-1}{k-1}$ then*

$$\text{m-sat}(n, \mathcal{D}) = \text{sat}(n, \mathcal{D}) = \binom{v}{k} + \left\lceil \frac{\lambda(n-v)}{k} \right\rceil. \tag{28}$$

Proof. Given $G \in \text{m-SAT}(n, \mathcal{D})$, let $V \subset V(G)$ consist of all vertices whose degree (that is, the number of containing it edges) is at most $\lambda - 1$. Clearly, V must span the complete k-graph, for otherwise the addition of a missing edge $E \in V^{(k)}$ to G cannot create any forbidden subgraph. Thus $e(G) \geq \min_{v \in [n]} f(v)$, where $f(v) = \binom{v}{k} + \frac{\lambda(n-v)}{k}$, which implies the lower bound on m-sat.

Conversely, let $n' = n - v$ and let G be the nearly λ-regular k-graph G' on $[n']$ built in Lemma 19, plus the complete k-graph on $V = [n'+1, n]$ and (if $D = \{x \in [n'] : d(x) < \lambda\} \neq \emptyset$) plus an edge E intersecting $[n']$ in the set D. (Note that $v \geq k - |D|$ if $D \neq \emptyset$: otherwise $G' + K^k(V) \in \text{m-SAT}(n, \mathcal{D})$ contradicts our lower bound.)

All vertices in $[n']$ have degree λ and any missing edge (which must intersect $[n']$) creates a forbidden subgraph. Also, G is $\mathcal{D}$-free: if $\binom{v-1}{k-1} \geq \lambda$, then we obtain the contradiction $f(v) \geq \lambda n/k > f(k-1)$. The required $G \in \text{SAT}(n, \mathcal{D})$ is built. ∎

Next, let $t = 2$ and $\mathcal{D} = \mathcal{D}(\lambda, 2, k)$, that is, we forbid $\lambda + 1$ edges having 2 common vertices. The *Turán number* $t(n, t, k) = \text{ex}(n, K_k^t)$ is the maximum size of a K_k^t-free t-graph of order n. Define $\alpha(n, t, k) = \binom{n}{t} - t(n, t, k)$. We are able to compute asymptotically $\text{sat}(n, \mathcal{D})$.

Theorem 21 *Given $\lambda \geq 2$ and $k \geq 3$, let $c = \lambda / \binom{k}{2}$. Then, for any $n \geq$* $\max(k + c - 1, kc^{1/(k-2)})$,

$$\text{m-sat}(n, D(\lambda, 2, k)) \geq c\alpha(n, 2, k). \tag{29}$$

On the other hand, for any fixed λ and k,

$$\text{sat}(n, \mathcal{D}(\lambda, 2, k)) \leq c\alpha(n, 2, k) + O\left(\frac{n^2}{(\log n)^{1/(\binom{k}{2}-1)}} \right). \tag{30}$$

Proof. Given a monotonically $\mathcal{D}$-saturated k-graph H, we build, on the same vertex set, the 2-graph G so that $\{i, j\} \in E(G)$ iff there are at least λ H-edges containing both $i, j \in V(H)$. Clearly, any k-set E independent in G must be an edge of H, for otherwise the addition of E to H does not create a forbidden subgraph. This implies that

$$e(H) \geq L(G) = k_k^2(\overline{G}) + ce(G), \tag{31}$$

where $k_k^2(\overline{G})$ denotes the number of K_k^2-subgraphs of $\overline{G}$, the complement of G. We want to find, for which 2-graphs G, the right-hand side of (31) is minimized. By a theorem of Bollobás [Bol76] (for some extensions see Schelp and Thomason [ST98]), this happens if $\overline{G}$ is a complete multipartite 2-graph (that is, if G is a disjoint union of complete graphs). If the parts are of sizes $n_1 \geq n_2 \geq \ldots \geq n_l$, then we have to minimize

$$L(G) = k_k^2(\overline{G}) + ce(G) = \sum_{A \in [l]^{(k)}} \prod_{i \in A} n_i + c \sum_{i=1}^{l} \binom{n_i}{2}, \tag{32}$$

given the condition $\sum_{i=1}^{l} n_i = n$.

Suppose that $l \geq k$. Let G' be obtained from G by merging the smallest two parts together. This adds $n_{l-1}n_l$ extra edges to G, but this eliminates all K_k^2-subgraphs of $\overline{G}$ intersecting both of the affected parts, that is,

$$k_k^2(\overline{G}) - k_k^2(\overline{G'}) = n_{l-1}n_l \sum_{A \in [l-2]^{(k-2)}} \prod_{i \in A} n_i. \tag{33}$$

We claim that $\sum_{A \in [l-2]^{(k-2)}} \prod_{i \in A} n_i \geq c$. As n_l and n_{l-1} are two smallest parts, it is enough to verify the inequality for $n_2 = \ldots = n_{l-2} = n_{l-1} = n_l = x$ in which case it reduces to to

$$g(x) = \binom{l-3}{k-2} x^{k-2} + \binom{l-3}{k-3} (n - (l-1)x) x^{k-3} \geq c. \tag{34}$$

Taking the derivative, one can see that the minimum of g over $x \in [1, n/l]$ is achieved either for $x = 1$ or for $x = n/l$. For $x = 1$, the right-hand side of (34) is $h(l) = \binom{l-3}{k-2} + \binom{l-3}{k-3}(n - l + 1)$ and, for any $l \in [k, n]$, the inequality $h(l) \geq c$ is true as $h(k) \geq c$ and

$$h(l+1) - h(l) = \tbinom{l-3}{k-3} + (n-l+1)\tbinom{l-3}{k-4} - \tbinom{l-2}{k-3} = (n-l)\tbinom{l-3}{k-4} \geq 0.$$

For $x = n/l$,

$$g(n/l) = \binom{l-2}{k-2} \left(\frac{n}{l}\right)^{k-2} = \frac{n^{k-2}}{(k-2)!} \prod_{i=1}^{k-2} \left(1 - \frac{i+1}{l}\right)$$

which is clearly minimized for $l = k$. But $g(n/k) \geq c$ by our assumptions.

Thus we may assume that $l \leq k-1$. But then $k_k^2(G) = 0$ and $e(G)$ is minimal if we have exactly $k-1$ parts of nearly equal sizes (i.e. if $\overline{G}$ is the Turán graph) and (29) follows.

To demonstrate the upper bound we have to use as bricks almost optimal sub-designs. Rödl [Röd85] was first to show that for fixed λ, k, t there exists a t-(v, k, λ)-sub-design with $\lambda\binom{v}{t}/\binom{k}{t} + o(v^t)$ edges, $v \to \infty$, that is, asymptotically approaching the absolute upper bound. The error term was made more specific by Gordon, Kuperberg, Patashnik and Spencer [GPKS96] who showed it to be $O(F(v))$, where $F(v) = v^t/(\log v)^{1/D}$ and $D = \binom{k}{t} - 1$. Gordon, Kuperberg and Patashnik [GKP95] present a few different methods suitable for practical construction of nearly optimal sub-designs.

Let us construct $G \in \mathrm{SAT}(n, \mathcal{D})$ showing that (29) is asymptotically correct. Partition $[n] = V_1 \cup \ldots \cup V_{k-1}$ into $k-1$ nearly equal parts. On each part V_i construct a maximum 2-$(|V_i|, k, \lambda)$-sub-design H_i. The union of $H_1, \ldots, H_{k-1}$ is obviously $\mathcal{D}$-free and has the size within $O(F(n))$ of (29). Completing it in an arbitrary way to $G \in \mathrm{SAT}(n, \mathcal{D})$, we add $O(F(v))$ extra edges as each extra edge intersect some part in at least 2 vertices while each H_i has $O(F(v))$ 2-sets covered by strictly less than λ edges. The theorem is proved. ∎

Finally, let us consider the general case $t \geq 3$. It seems that $\mathrm{sat}(n, \mathcal{D}(\lambda, t, k))$ is generally related to $\alpha(n, t, k)$.

Theorem 22 *For any fixed λ, t and k,*

$$\text{m-sat}(n, \mathcal{D}(\lambda, t, k)) \geq (1 - o(1))\lambda\alpha(n, t, k)\binom{k}{t}^{-1}, \tag{35}$$

as n tends to infinity.

Proof. Let $H \in \text{m-SAT}(n, \mathcal{D}(\lambda, t, k))$. Let the t-graph G consist of all t-sets covered by at least λ edges of H. Similarly to the above, we note that any k-subset of $[n]$ not spanning an edge in G, must belong to $E(H)$ and therefore,

$$e(H) \geq \lambda e(G)\binom{k}{t}^{-1} + k_k^t(\overline{G}). \tag{36}$$

If $e(\overline{G}) \leq (1 + o(1))t(n, t, k)$ then the first summand in the right-hand side of (36) itself gives the desired lower bound. Otherwise, the result of Erdős and

Simonovits [ES83] implies that the second summand is $\Theta(n^k)$ which is far more than required. ∎

We do not have many structural results related to the Turán problem for complete hypergraphs. Sidorenko [Sid95] mentions the following conjectures.

$$\alpha(n,3,k) = \left(\frac{2}{k-1}\right)^2 \binom{n}{3} + o(n^3), \tag{37}$$

$$\alpha(n,4,5) = \frac{5}{16}\binom{n}{4} + o(n^4). \tag{38}$$

Recall that $\alpha(n,t,k) = \binom{n}{t} - t(n,t,k)$ is the minimum size of an $\alpha(n,t,k)$-*graph*, that is, a t-graph on n vertices in which any k-set spans at least one edge.

Example 23 *Let $\mathcal{D} = \mathcal{D}(\lambda,3,k)$, where either $k=4$ or $k \geq 5$ is odd. Then there is a $\mathcal{D}$-saturated k-graph H with $\lambda(\frac{2}{k-1})^2\binom{n}{3}/\binom{k}{3} + o(n^3)$ edges. In particular, if* (37) *is true, then H is asymptotically extremal.*

Proof. Let $k=4$. Let $m = \lfloor n/3 \rfloor$. Define $A_i = [(i-1)m+1, im]$, $i \in [3]$. The graph G on $[3m]$ consisting of all triples $\{x,y,z\}$ with $x,y \in A_i$ and $z \in A_i \cup A_{i+1}$, where $A_4 = A_1$, is an $\alpha(3m,3,4)$-graph with approximately $\frac{4}{9}\binom{n}{3}$ edges.

Consider the graph H' consisting of edges $E = \{w,x,y,z\}$ with $\{x,y,z\} \in A_i^{(3)}$ and $w \in A_{i+1}$ (then $E^{(3)} \subset E(G)$), $i \in [3]$, such that $u+x+y+z$ is congruent modulo m to an element in $[\lambda]$. Let $D \in E(G)$. For example, suppose that D consists of $x,y \in A_1$ and $w \in A_2$. To find z with $\{w,x,y,z\} \in E(H')$ we have to satisfy $w+x+y+z \equiv j \pmod{m}$ for some $j \in [\lambda]$; there are λ solutions, but we may have to discard possible degenerate cases when $z = x$ or $z = y$. A similar claim is true if $D \subset A_i$. Hence, each G-edge, except $O(n^2)$ edges, is covered by exactly λ edges of H'.

It is therefore clear that if we complete the $\mathcal{D}$-free graph H' to any $\mathcal{D}$-saturated graph H on $[n]$, then we add only $O(n^2)$ edges; therefore, H has the required size.

For $k = 2l+1$, $l \geq 2$, an example of an $\alpha(n,3,k)$-graph G attaining (37) is obtained by partitioning $[n] = A_1 \cup \ldots \cup A_l$ into nearly equal parts and letting $G = \cup_{i\in[l]} K^3(A_i)$. The result of Rödl [Röd85] implies that we can find a $\mathcal{D}$-free k-graph on each A_i which is a nearly optimal 3-$(|A_i|,k,\lambda)$-sub-design; let H' be the union of these. Completing it arbitrarily to a $\mathcal{D}(\lambda,3,k)$-saturated graph, we add only $o(n^3)$-extra edges, which proves the claim. ∎

However, we do not know any matching construction for $t = 3$ and even $k \geq 6$. In this case, a conjectured extremal $\alpha(n,t,k)$-graph G is the disjoint union of complete 3-graphs plus at least one $\alpha(m,3,4)$-extremal graph. The last graph causes us the problem: the constructions by Kostochka [Kos82] do not admit an almost perfect covering by k-edges, $k \geq 6$.

Here is a short explanation why. In all Kostochka's graphs we have three equisized sets $A_1 \cup A_2 \cup A_3 = [n]$ and let $G = (\cup_{i=1}^3 K^3(A_i)) \cup G'$, where G' consists of $\frac{1}{3}\binom{n}{3} + o(n^3)$ other edges. Also, any k-set E with $E^{(3)} \subset E(G)$ has the property that $|K \cap A_i| \geq k-2$ for some i, so it can cover at most $l = \binom{k}{3} - \binom{k-2}{3}$ edges of G'. Hence, we need at least $e(G')/l$ covering edges, which exceeds $(\frac{2}{k-1})^2\binom{n}{3}/\binom{k}{3} + o(n^3)$ for $k \geq 8$. (For $k = 6$ we need a slightly more refined argument.)

For similar reasons, there is no almost perfect covering of the construction by de Caen, Kreher and Wiseman [dCKW88] which gives the upper bound in (38).

Unfortunately, we do not know any other, essentially different, constructions attaining (37) or (38) and we do not have any likely guess what $\mathrm{sat}(n,\mathcal{D})$ could be then.

6.3 Triangular Families

The notion of a triangle-free 2-graph can be extended to hypergraphs in the following way: a k-graph is *triangle-free* if the symmetric difference of any two distinct edges is not contained in a third edge. Clearly, this is the same as forbidding the *triangular family* $\mathcal{T}_k$ which consists of all k-graphs with three edges E_1, E_2, E_3 such that $E_1 \triangle E_2 \subset E_3$.

Katona [Kat74] raised the problem of computing $\mathrm{ex}(n,\mathcal{T}_3)$ which was solved by Bollobás [Bol74] who showed that the complete 3-partite 3-graph with parts of nearly equal sizes is a maximum triangle-free 3-graph. Bollobás [Bol74] stated the general conjecture that the analogous construction gives $\mathrm{ex}(n,\mathcal{T}_k)$ for any $k \geq 4$; Sidorenko [Sid87] proved that this is the case for $k = 4$.

Concerning the sat-function, we have the following obvious example of a $\mathcal{T}_k$-saturated graph: the *pyramid* $P = P(k-1, n-k+1; k-1, 1)$ which consists of all k-subsets of $[n]$ containing the set $[k-1]$ called *basic*. Indeed, any missing edge E intersects $[k,n]$ in at least 2 points and creates a forbidden subgraph on the set $E \cup [k-1]$. Thus

$$\mathrm{sat}(n, \mathcal{T}_k) \le n - k + 1, \quad n \ge k + 1$$

and this might be sharp. It is remarkable that P can be viewed as the complete k-partite k-graph with $k-1$ parts consisting of only one vertex.

In the general case we are able to prove only the following.

Theorem 24 *Let $k \ge 3$ be fixed. Then*

$$n - O(\log n) \le \mathrm{sat}(n, \mathcal{T}_k) \le n - k + 1.$$

Proof. We have to prove the lower bound. Let G be a minimum $\mathcal{T}_k$-saturated graph on $[n]$; $e(G) \le n - k + 1$. Consecutively choose $G_1, G_2, \ldots \subset G$ as follows: let e_{j+1} be the largest integer such that the k-graph H_j, $E(H_j) = E(G) \setminus (E(G_1) \cup \ldots \cup E(G_j))$, contains a $P(k-1, e_{j+1}; k-1, 1)$-subgraph and let G_{j+1} be any such subgraph. We terminate the procedure when $b_j = n - e_{[j]} - j(k-1)$ is less than $\max(j, k)$. (We denote $e_{[j]} = \sum_{i \in [j]} e_i$, etc.)

Let $j \ge 0$ and suppose we have chosen $G_1, \ldots, G_j$. Let B_j consist of some b_j vertices not covered by an edge of G_i, $i \in [j]$; B_j exists as $v(G_i) = e_i + k - 1$. (We let $b_0 = n$.) Label all $(k-1)$-subsets of $[n]$ by $A_1, \ldots, A_l$, $l = \binom{n}{k-1}$. Let d_i be the number of edges of H_j containing A_i, $i \in [l]$. Clearly,

$$d_{[l]} = ke(H_j) \le k(n - k + 1 - e_{[j]}) = k(b_j + (j-1)(k-1)) < k^2 b_j. \tag{39}$$

The number of ways to add an element of $B_j^{(k)}$ creating a forbidden subgraph with any given $E_1, E_2 \in [n]^{(k)}$ is at most $\binom{b_j - 2}{k-2} + O(1)$ if $|E_1 \cap E_2| = k - 1$ and it is $O(b_j^{k-4})$ otherwise. As the addition of any $E \in B_j^{(k)} \setminus E(H_j)$ to H_j creates a forbidden subgraph (because E is disjoint from any edge of G_i, $i \in [j]$), we conclude that

$$O(b_j^{k-4}) \binom{e(H_j)}{2} + \binom{b_j - 2}{k-2} \sum_{i \in [l]} \binom{d_i}{2} \ge \binom{b_j}{k} - e(H_j), \tag{40}$$

and, by (39),

$$\sum_{i \in [l]} \binom{d_i}{2} \ge \frac{b_j^2}{k(k-1)} - O(b_j). \tag{41}$$

We have $e_{j+1} = \max_{i \in [l]} d_i$. The convexity of the $\binom{x}{2}$-function implies that the left-hand side of (41) does not exceed $\frac{d_{[l]}}{e_{j+1}} \binom{e_{j+1}}{2} < \frac{1}{2} k^2 b_j e_{j+1}$. Therefore, we obtain that

$$e_{j+1} \geq \frac{2b_j}{k^3(k-1)} - O(1).$$

From this inequality (and from the fact that $e_{j+1} \geq 1$ if $b_j \geq k$) we deduce the following inequality

$$b_{j+1} \leq \min\left(\left(1 - \tfrac{2}{k^3(k-1)}\right) b_j + O(1),\ b_j - k\right). \tag{42}$$

It is clear that, starting with $b_0 = n$, we stop after $j = O(\log n)$ steps. Now,

$$e(G) \geq e_{[j]} = n - b_j - j(k-1) = n - O(\log n).$$

The theorem is proved. ∎

Let us consider the case $k = 3$; note that $\mathcal{T}_3$ contains only 2 non-isomorphic graphs, S_4^3 and T_3:

$$\begin{aligned} E(S_4^3) &= \{\ \{1,2,3\},\ \{1,2,4\},\ \{1,3,4\}\ \},\\ E(T_3) &= \{\ \{1,2,3\},\ \{1,2,4\},\ \{3,4,5\}\ \}. \end{aligned}$$

Theorem 25 *For any $n \geq 4$, $\mathrm{sat}(n, \mathcal{T}_3) = n - 2$.*

Proof. Let G be any $\mathcal{T}_3$-saturated graph on $[n]$. Make a list of all edges of G and, consecutively and as long as possible, *merge* together any two sets in the list sharing at least 2 vertices (that is, replace then by their union.) Call the resulting sets $C_1, \ldots, C_l \subset [n]$ *components.* Let $v_i = |C_i|$. Define the 2-graph H on $[n]$ by

$$E(H) = \{\{x,y\} \in [n]^{(2)} : \{x,y\} = E_1 \triangle E_2 \text{ for some } E_1, E_2 \in E(G)\}.$$

Consider any component C. It is easy to see by induction on $|C|$ that C is composed of at least $|C| - 2$ edges of G.

Note that if $E \in E(H[C])$ then any $E_1, E_2 \in E(G)$ with $E_1 \triangle E_2 = E$ share 2 vertices and so belong to the same component C'; but $E \subset C' \cap C$ so necessarily $C' = C$.

Claim 1 For every component C, $\Delta(H[C]) \leq e(G[C]) - 1$.

Let $x \in C$ be arbitrary. For each $\{x, y\} \in E(H[C])$, choose $D_y, E_y \in E(G)$ with $D_y \triangle E_y = \{x, y\}$ and $E_y \ni y$. If $\{x, z\}$ is another edge of $H[C]$ then $E_y \neq E_z$: indeed, otherwise $D_z \triangle E_z = \{x, z\} \subset D_y$ and G contains a forbidden subgraph. Hence, $d(x) \leq e(G[C]) - 1$ (we must have at one G-edge incident to x) and the claim is proved.

Claim 2 If $e(G[C]) \leq |C| - 1$ then for any $x \in [n] \setminus C$ there is a component $C' \ni x$ intersecting C.

By Claim 1, there exists $\{a, b\} \in E(\overline{H}[C])$. As $x \notin C$, $E = \{a, b, x\} \notin E(G)$. Consider a forbidden subgraph F created by E. We are home if $\{a, x\}$ or $\{b, x\}$ is covered by E_1 or E_2, where $E(F) = \{E, E_1, E_2\}$. If $\{a, b, y\} \in E(F)$ then $y \in C$ and the remaining edge of F contains both x and y. Finally, if $E_1 \triangle E_2 \subset E$ then, as $\{a, b\} \notin E(H)$, x belongs to the component containing E_1 and E_2 which is the required one.

The claim is proved. In particular, $C_{[l]} = V(G)$.

Now, if every component C spans at least $|C| - 1$ edges then we are home: by Claim 2 relabel components $C_1, \ldots, C_l$ so that $C_i \cap C_{[i-1]} \neq \emptyset$, $i \in [2, l]$, and it is easy to show by induction on i that $C_{[i]}$ is made of at least $|C_{[i]}| - 1$ edges, which gives $e(G) \geq n - 1$.

So, suppose that, for example, $e(G[C_1]) = |C_1| - 2$. If for every $x \in V(G) \setminus C_1$, there are two distinct components containing x and intersecting C_1 then are home:

$$\begin{aligned} e(G) \; &\geq \; \sum_{i \in [l]} (v_i - 2) \; = \; v_1 - l - 1 + \sum_{i \in [2,l]} (v_i - 1) \\ &\geq \; v_1 - l - 1 + \max(2l - 2, 2(n - v_1)) \; \geq \; n - 2. \end{aligned} \tag{43}$$

So let C_2 be the only component containing some vertex $x \notin C_1$ and intersecting C_1. Let $\{y\} = C_1 \cap C_2$. Let $z \in V(G) \setminus C_{[2]}$ be arbitrary. (The below argument works without any changes if $C_1 \cup C_2 = V(G)$.)

If $\{x, z\} \subset C_i$, for some $i \in [3, l]$, then, by the choice of x, $C_i \cap C_1 = \emptyset$ and, by Claim 2, there exists another component through z intersecting C_1.

If no component contains both x and z then, for every $y' \in C_1 \setminus \{y\}$, $E = \{x, y', z\} \notin E(G)$ and considering a forbidden subgraph created by E we conclude that, for some $i \in [3, l]$, $\{y', z\} \subset C_i$ (as $\{x, y'\}$ cannot lie within a component by the definition of x). As $|C_1| \geq 3$, we have at least 2 distinct components containing z and intersecting C_1.

Now the argument similar to (43) shows that $C_{[3,l]}$ is made of at least $n - |C_1 \cup C_2|$ edges, which gives $e(G) \geq n - 3$.

Can we have $e(G) = n - 3$? If we have the equality then every C_i, $i \in [3, l]$, must intersect $C_1 \cup C_2$ in exactly one vertex and $e(G[C_j]) = |C_j| - 2$, $j \in [l]$. By Claim 1, there exists $y_i \in C_i$ such that $\{y, y_i\} \notin E(H)$, $i = 1, 2$. But then $\{y, y_1, y_2\} \notin E(G)$ (e.g. because it intersects C_1 in two vertices) and the consideration of a created forbidden graph yields a component containing both y_1 and y_2. Hence, $e(G) > n - 3$ as required. ∎

Remark. Our further analysis has not yet yielded any characterization of the cases of equality: we have got stuck considering different cases and, even if we had succeeded, the proof would have been rather long. Therefore, we present only some other constructions which we have discovered in our search. First, there is another minimum $\mathcal{T}_3$-saturated graph of order 7: let $V(G) = [7]$ and

$$E(G) = \{\ \{1, 2, 5\},\ \{1, 3, 6\},\ \{1, 4, 7\},\ \{2, 3, 4\},\ \{5, 6, 7\}\ \}.$$

Also, concerning the m-sat-function, we have yet another construction with $n-2$ edges for any $n \geq 6$: add, to the pyramid $P(2, n-4; 2, 1)$ with basic vertices a, b, new vertices x, y and new edges $\{x, y, a\}$ and $\{x, y, b\}$.

6.4 K_m-Saturated Graphs

Duffus and Hanson [DH86] consider $\mathrm{sat}(n, K_m, l)$ which is the minimum size of

$$G \in \mathrm{SAT}(n, K_m, l) = \{G \in \mathrm{SAT}(n, K_m) : \delta(G) \geq l\}.$$

Of course, any K_m-saturated graph G has minimal degree at least $m - 2$, so we assume $l \geq m - 1$.

Duffus and Hanson [DH86] proved that, for $n \geq 5$, $\mathrm{sat}(n, K_3, 2) = 2n - 5$ and, for $n \geq 10$, $\mathrm{sat}(n, K_3, 3) = 3n - 15$. However, their general lower bound [DH86, Theorem 2], which states that $\mathrm{sat}(n, K_m, l) \geq \frac{l+m-2}{2} n + O(1)$, is far from the actual value. Trying to improve this bound, we showed that $\mathrm{sat}(n, K_m, l) = ln + O(\frac{n \log\log n}{\log n})$ for any fixed $l \geq m - 1$. Later, we learned that Alon, Erdős, Holzman and Krivelevich [AEHK96, Theorem 2] showed that any $G \in \mathrm{SAT}(n, K_m)$ with

$O(n)$ edges has an independent set of size $n - O(\frac{n}{\log\log n})$, which implies that $\mathrm{sat}(n, K_m, l) = ln + O(\frac{n}{\log\log n})$. However, we decided to present our proof because it improves all these bounds and we think that our general Theorem 26 is of independent interest.

However, the question of Bollobás [Bol95, p. 1271] whether $\mathrm{sat}(n, K_3, l) = ln + O(1)$ for any fixed $l \geq 4$, remains open.

Let us give a construction of $G \in \mathrm{SAT}(n, K_m, l)$ with $ln + O(1)$ edges: take $G = K_{m-3} + K_{l-m+3,n-l}$ which has minimal degree l for $n \geq 2l - m + 3$. The complete bipartite graph $K_{l-m+3,n-l}$ does not contain a triangle but the addition of any new edge violates this; hence, G is K_m-saturated.

To prove our lower bound we need some preliminaries. Given any d, define $a_{d-m+2} = 2$ and, consecutively for $j = d - m + 1, d - m, \ldots, 1, 0$,

$$\begin{aligned} c_{j+1} &= (m-2)(a_{j+1} - 1) + 1 \\ b_{j+1} &= (m-2)(c_{j+1} - 1) + 1 \\ b'_{j+1} &= \binom{d-j-1}{m-2}(b_{j+1} - 1) + 1, \\ a_j &= \binom{d-j-1}{m-2}(b'_{j+1} - 1) + 2. \end{aligned}$$

Finally, let $a = (1 + 2(d-1) + 2(d-1)^2)a_0$.

Given a K_m-saturated graph G, let A denote the set of G-edges connecting two vertices of degree at most d in G:

$$A = \{\{x, y\} \in E(G) : d(x) \leq d,\ d(y) \leq d\}.$$

The following theorem states that the size of A is bounded by $a = a(d, m)$ which does not depend on n. Note that we do not impose any restriction on the minimal degree of G.

Theorem 26 *For any $G \in \mathrm{SAT}(n, K_m)$, $m \geq 3$, we have $|A| < a$.*

Proof. Suppose, on the contrary, that $|A| \geq a$.

We prove, by induction on $j = 0, 1, \ldots, d - m + 2$, that we can find the following configuration in G: a_j-sets X_j and Y_j and j-sets U_j and V_j (all disjoint) such that (i) $X_j \cup Y_j$ induces in G exactly a_j edges which form a perfect matching between X and Y and belong to A; (ii) $\Gamma_{U_j \cup V_j}(x) = U_j$ for any $x \in X_j$ and $\Gamma_{U_j \cup V_j}(y) = V_j$ for any $y \in Y_j$.

For $j = 0$ (when U_0 and V_0 are empty), we take, one by one, edges from A. Once we have selected an edge $E \in A$, cross out all incident to E edges (at most $2(d-1)$ edges) and their neighbouring edges (of which at most $2(d-1)^2$ can belong to A). Hence, we can build an induced matching of size at least $|A|/(1+2(d-1)+2(d-1)^2) \geq a_0$ as required.

Suppose that $j \in [0, d-m+1]$ and we have X_j, etc., constructed. Choose $x \in X_j$; it has already got $j+1$ neighbours in G: the neighbour $y \in Y_j$ plus all j vertices of U_j. Let N_x denote the remaining neighbours of x; thus $|N_x| \leq d-j-1$. For any $z \in Y_j$ distinct from y, the addition of the edge $\{x, z\}$ must create a copy of K_m, say on a set $D_z \cup \{x, z\}$. Now, $D_z \subset \Gamma(x) \cap \Gamma(z) \subset N_x$.

Thus some set D_z, $z \in Y_j \setminus \{y\}$, appears at least $b'_{j+1} = \lceil (a_j - 1)/\binom{d-j-1}{m-2} \rceil$ times; suppose it is $D \in N_x^{(m-2)}$ which equals D_z for $z \in B' \subset Y_j \setminus \{y\}$, $|B'| = b'_{j+1}$. In a similar manner, we try to connect y to the X_j-matches of B'-vertices and find a set $E \in N_y^{(m-2)}$ spanning the complete graph and connected to every z from a set $B \subset X_j$ matched into B' of cardinality $b_{j+1} = \lceil b'_{j+1}/\binom{d-j-1}{m-2} \rceil$.

Clearly, no $z \in B$ can be connected to every vertex of D; otherwise D, z and the match of z in B' span K_m. Therefore, some $v \in D$ is not connected to at least $c_{j+1} = \lceil \frac{b_{j+1}}{m-2} \rceil$ vertices of B; let $C \subset B$ consist of all such vertices. Similarly, we can find $u \in E$, not connected to an a_{j+1}-set Y_{j+1} matched into C. Of course, $u \neq v$. Now, let $U_{j+1} = U_j \cup \{u\}$, $V_{j+1} = V_j \cup \{v\}$, and let $X_{j+1} \subset X_j$ consist of the matches of Y_{j+1}, which completes our induction.

At the end, we try to apply our argument again, for $j = d-m+2$. We obtain that $x \in X_j$ has at least $1+j+(m-2) > d$ neighbours, which contradicts the fact that $\{x, y\} \in A$, where y is the Y_j-match of x. ∎

Now we are ready to improve the result of Alon et al [AEHK96, Theorem 2] mentioned above. Let $\alpha(G)$ denote the maximal size of independent $Y \subset V(G)$.

Lemma 27 *For any $G \in \mathrm{SAT}(n, K_m)$ with $O(n)$ edges, we have*

$$\alpha(G) = n - O\left(\frac{n \log \log n}{\log n}\right).$$

Proof. Suppose $e(G) \leq Cn$. Let $d = \frac{\varepsilon \log n}{\log \log n}$ for some fixed $\varepsilon > 0$ and let $X = \{x \in V(G) : d(x) > d\}$. Now, $d|X|/2 \leq e(G) \leq Cn$ implies that

$$|X| \leq \frac{2Cn \log \log n}{\varepsilon \log n}.$$

By Theorem 26, $Y = V(G)\setminus X$ spans at most $a \leq n^{2\varepsilon(m-2)+o(1)}$ edges. Removing at most a vertices we can make Y independent; it has the required size if $\varepsilon < \frac{1}{2(m-2)}$. ∎

Clearly, $e(G) \geq \alpha(G)\delta(G)$. Therefore, Lemma 27 implies the following result.

Theorem 28 *For any fixed $l \geq m-1$, $\mathrm{sat}(n, K_m, l) = ln + O(\frac{n \log\log n}{\log n})$.* ∎

Part II

Weakly Saturated Hypergraphs

7 Introduction

In this part we move to studying weakly-saturated graphs. They are briefly mentioned in Section 1 which also contains an example how such a notion can naturally appear in real-life problems.

Let us give some basic definitions, describe what is known about the w-sat-function, and indicate which new results are obtained in this part.

7.1 Definitions

Let $\mathcal{F}$ be a family of forbidden r-graphs. An r-graph G of order 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 at least one new forbidden subgraph appears. Such an ordering of $E(\overline{G})$ is called *$\mathcal{F}$-proper*. Equivalently, $G \in \text{w-SAT}(n, \mathcal{F})$ if the *weak closure* $\text{Cl}_{\mathcal{F}}(G)$ is the complete r-graph on $V(G)$. (The *weak closure* is obtained by taking the iterated (strong) $\mathcal{F}$-closure (defined in Subsection 4.4) until it stabilizes: $\text{Cl}_{\mathcal{F}}(G) = \text{Cl}^*_{\mathcal{F}}(...(\text{Cl}^*_{\mathcal{F}}(G))...)$.) We are generally interested in

$$\text{w-sat}(n, \mathcal{F}) = \min\{e(G) : G \in \text{w-SAT}(n, \mathcal{F})\}.$$

Note that we do not require that G is $\mathcal{F}$-admissible as this does not affect $\text{w-sat}(n, \mathcal{F})$: if $G \in \text{w-SAT}(n, \mathcal{F})$ contains a forbidden subgraph $F \subset G$, then the graph obtained from G by the removal of any F-edge is still weakly $\mathcal{F}$-saturated, so G cannot be minimal. Clearly, $\text{w-sat}(n, \mathcal{F}) \leq \text{m-sat}(n, \mathcal{F}) \leq \text{sat}(n, \mathcal{F})$.

If the forbidden family consists of only one member, $\mathcal{F} = \{F\}$, then we use the shortcuts $\text{w-SAT}(n, F) = \text{w-SAT}(n, \mathcal{F})$, etc.

7.2 Survey

Let us give a short survey of w-sat-type results. Unfortunately, not much is known about the w-sat-function.

Tuza [Tuz92] showed that, for any fixed r-graph F,

$$\text{w-sat}(n,F) = \Theta(n^{s(F)}). \tag{44}$$

Here $s(F) \in [0, r-1]$ is what he calls the *local sparseness* of F:

$$s(F) = \min\{s(E) : E \in E(F)\}, \tag{45}$$

where the *sparseness* of an edge $E \in E(F)$ is the smallest natural number s for which there is an $A \subset E$ with $|A| = s+1$ such that $A \subset E' \in E(F)$ implies $E' = E$.

Alon [Alo85] proved that, for any fixed 2-graph F, the ratio $\text{w-sat}(n,F)/n$ tends to a limit as $n \to \infty$.

Apparently, w-sat-type problems were first considered by Bollobás [Bol67c] who made a conjecture about the value of $\text{w-sat}(n, K_m^2)$. This conjecture was proved by Frankl [Fra82] and by Kalai [Kal84, Kal85]; the result is implicit in Lovász [Lov77]; see also Alon [Alo85]. They proved that

$$\text{w-sat}(n, K_m^r) = \binom{n}{r} - \binom{n-m+k}{r}, \quad n \ge m > k. \tag{46}$$

In fact, Alon [Alo85] proved a more general result: he computed the w-sat-function for $K_{m_1}^{r_1} \otimes \ldots \otimes K_{m_t}^{r_t}$, where $\otimes$ denotes the *join* operator defined in Section 12. (A different proof of Alon's result is presented by Yu [Yu93].)

Kalai [Kal85] showed that, for the complete bipartite graph K_{st},

$$\text{w-sat}(n, K_{st}) \ge (s-1)n - \binom{s-1}{2}, \quad 2 \le s \le t, \tag{47}$$

which is sharp for $s = t$ and $n \ge 3s-2$.

Kalai [Kal85] also proved that, for the wheel $W_m = v + C_m$, we have

$$\text{w-sat}(n, W_m) \ge 2n - 3, \tag{48}$$

while it is easy to show that $\text{sat}(n, W_m) \le 2n - 3 + \varepsilon$, where $\varepsilon = 0$ if $m \le n-2$ is odd and $\varepsilon = 1$ if m is even or if $m = n-1$, cf. Theorems 49–51 and Lemma 59.

Tuza [Tuz88, Conjecture 7] conjectured that

$$\text{w-sat}\,(n, \mathcal{H}_r(r+1, l)) = \binom{n-r+l-2}{l-2} \quad n \ge r+1 \ge l \ge 3, \tag{49}$$

where the *uniform family* $\mathcal{H}_r(m,l)$ consists of all r-graphs of order m and size l. Clearly, $\mathcal{H}_r(r+1,r+1) = \{K^r_{r+1}\}$, so (46) implies (49) for $l = r+1$. The case $l = 3$ of Tuza's conjecture was settled by Erdős, Füredi and Tuza [EFT91].

These were perhaps all known results on $\text{w-sat}(n,\mathcal{F})$ for non-trivial specific instances of $\mathcal{F}$.

7.3 Our Approach

The characteristic feature of w-sat-type problems is that, given a particular forbidden family $\mathcal{F}$, it is usually fairly easy to come up with a correct example of $G_n \in \text{w-SAT}(n,\mathcal{F})$, which gives us an upper bound on $\text{w-sat}(n,\mathcal{F})$. (And, as a rule, we have many different extremal graphs.) However, it is usually very hard to prove the matching lower bound. So, techniques for establishing lower bounds are of importance.

The notion of dependence turned out to be useful; for example, all proofs of (46) exploit some form of it. This approach was most clearly formulated by Kalai [Kal85]: if we have a matroid $\mathcal{M}$ on $[n]^{(r)}$ such that any $F \in \mathcal{F}$ is an $\mathcal{M}$-chain, then

$$\text{w-sat}(n,\mathcal{F}) \geq R_{\mathcal{M}}([n]^{(r)}), \tag{50}$$

the rank of $\mathcal{M}$. (An r-graph F is an *$\mathcal{M}$-chain* if, for any embedding $V(F) \subset [n]$, any edge $E \in E(F)$ lies in the $\mathcal{M}$-span of $E(F) \setminus \{E\}$.) See Lemma 33 for a proof of (50).

We base our approach (which is described in detail in Section 9) on this idea; we exploit what we call *gross* and *count matroids.*

Gross matroids are constructed by means of exterior algebra. They were considered by Kalai [Kal90] (but for other purposes); we define them in Subsection 8.1. In brief, the gross matroid $\mathcal{G}_G$ of an r-graph G is a matroid on r-uniform set systems with G being a base; thus its rank is $e(G)$. Now, if every $F \in \mathcal{F}$ is a $\mathcal{G}_G$-chain, then

$$\text{w-sat}(n,\mathcal{F}) \geq R_{\mathcal{G}_G}([n]^{(r)}). \tag{51}$$

The lower bound (51) is said to be *g-proved.* If this method gives the actual value of $\text{w-sat}(n,\mathcal{F})$, then we say that $\mathcal{F}$ admits a *g-proof* for n. A related method (*g'-proof*) is also introduced.

The principal difficulty of the matroid approach (50) is that it is not clear at all how to search for a suitable matroid $\mathcal{M}$. However, if we have $G \in$ w-SAT$(n, \mathcal{F})$ conjectured to be minimal, then $\mathcal{G}_G$ is a good candidate for $\mathcal{M}$. If each forbidden graph is a $\mathcal{G}_G$-chain, then, by (51), we know w-sat$(n, \mathcal{F})$ exactly. In this case we say that the pair $(\mathcal{F}, G)$ admits a *G-proof.*

Our *count matroid* is a general and natural extension of the construction by White and Whiteley [WW84], see Subsection 8.2. For example, our count matroids admit many polynomials in n as the rank function while the original definition yielded linear functions only. If $\mathcal{M}$ in (50) is a count matroid, then the lower bound (50) is said to be *c-proved.* If the bound is sharp, then $\mathcal{F}$ admits a *c-proof* for n. Here as well, if we have a conjecture on w-sat$(n, \mathcal{F})$, then there is one particular count matroid which is worth looking at; if this method works, then we have a *C-proof.*

Unfortunately, our approaches do not always succeed: we can indicate many concrete pairs $(\mathcal{F}, G)$ not admitting a C/G-proof with $G \in$ w-SAT$(n, \mathcal{F})$ being minimal. However, using these techniques we have managed to prove many new results which we are going to describe now.

Given sequences of integers $\mathbf{s} = (s_1, \ldots, s_t)$ and $\mathbf{r} = (r_1, \ldots, r_t)$, the *pyramid* $P(\mathbf{s}; \mathbf{r})$ is the r-graph, $r = r_1 + \ldots + r_t$, with vertex set being the disjoint union $S_1 \cup \ldots \cup S_t$, $|S_i| = s_i$, and with the edge set consisting of those r-subsets which, for every $i \in [t]$, intersect $S_1 \cup \ldots \cup S_i$ in at least $r_1 + \ldots + r_i$ vertices. The notion of a pyramid is rather general: we obtain, as partial cases,

$$\begin{aligned} K_m^r &= P(m; r), \\ S_m^r &= P(1, m-1; 1, r), \\ K_l + E_m &= P(l, m; 1, 1), \\ \mathcal{H}_r(r+1, l) &= P(r-l+1, l; r-l+1, l-1), \end{aligned}$$

and more. Instances of pyramids appear explicitly quite often in the literature.

Applying gross matroids, we compute w-sat$(n, P(\mathbf{s}; \mathbf{r}))$ for *all feasible* sets of parameters n, $\mathbf{s}$ and $\mathbf{r}$, see Subsection 10.1. Among other things, this implies (46) and computes w-sat$(n, \mathcal{H}_r(r+1, l))$, confirming the formula (49) conjectured by Tuza [Tuz88, Conjecture 7].

Erdős, Füredi and Tuza [EFT91] asked for a description of all minimum weakly $\mathcal{H}_r(r+1, 3)$-saturated graphs. In general, G/g/g′-proofs do not provide any good characterization of the cases of equality, but our Theorem 44 does this for $\mathcal{H}_r(r+1,3)$ by providing a different (combinatorial) proof which employs some ideas from [EFT91]. (In fact, $\mathcal{H}_r(r+1,3)$ admits a C-proof.) In Section 10.2 we provide a construction of $G \in$ w-SAT$(n, \mathcal{H}_r(m,l))$, for all n, k, l and m, which we conjecture to be minimal. Applying count matroids, we determine more values of w-sat$(n, \mathcal{H}_r(m,l))$. Applying the g$'$-proof technique, we compute exactly w-sat$(n, \mathcal{H}_2(m,l))$ for all possible n, m and l and obtain some asymptotic results. Also, we observe that we have incidentally computed (with a g$'$-proof) the w-sat-function for any initial segment of $[n]^{(2)}$ in the colex order.

Our more general results of Section 11 imply in particular that if (F, G) admits a G/g/g$'$-proof and every $r-1$ vertices of F are covered by an edge, then the pair $(\mathrm{cn}(F), \mathrm{cn}(G))$ admits a G/g/g$'$-proof. (The *cone* $\mathrm{cn}(F)$ of an k-graph F is obtained by adding to F a new vertex v and all r-edges containing v.)

In the class of 2-graphs, for example, we have $\mathrm{cn}^l(F) = K_l + F$. The following 2-graphs are shown to admit a G/g/g$'$-proof: complete graphs, stars, odd cycles, initial colex-segments of $[n]^{(2)}$, disjoint edges, paths (more generally, almost every forest or tree), and some others; please refer to Subsection 10.3 for details. Therefore, we are able to compute the w-sat-function for $K_l + F$, where F is any of these graphs.

Note that $\mathrm{cn}(K_m^r) = K_{m+1}^r$ and K_r^r, the single edge, trivially admits a G-proof as w-sat$(n, K_r^r) = 0$. This shows that complete graphs admit a G-proof and gives another proof of (46).

In Section 12 we define the $\otimes$-operator, which we call *join*, and prove among other things that if every pair $(\mathcal{F}_i, G_i)$, $i \in [t]$, admits a G/g-proof, then so does the pair $(\mathcal{F}_1 \otimes \ldots \otimes \mathcal{F}_t, \mathbf{G})$, where $\overline{\mathbf{G}} = \overline{G_1} \otimes \ldots \otimes \overline{G_t}$. As complete graphs admit a G-proof, the computation of the w-sat-function for joins of complete graphs by Alon [Alo85] (another proof is presented by Yu [Yu93]) is a special instance of our result. By applying the join operator, we can indicate many new graphs for which we can compute the w-sat-function exactly.

8 Matroids

Here we define gross and count matroids and establish their basic properties. (For an introduction to matroid theory, we refer the reader to the texts by Welsh [Wel76] or Oxley [Oxl92].) Our approach to w-sat-type problems, which exploits these notions, is described in Section 9.

8.1 Gross Matroids

Here we define the notion of a *gross matroid* by means of exterior algebra. Some background in multilinear algebra is included; for a more comprehensive treatment of the topic, the reader may consult Bourbaki [Bou74] or Marcus [Mar75].

8.1.1 Exterior Algebra

Let V be an n-dimensional real vector space with a basis $\mathbf{e} = \{e_1, \ldots, e_n\}$. Its *exterior algebra* $\bigwedge V$ is the 2^n-dimensional vector space with the formal basis $(e_A)_{A \subset [n]}$. (We identify e_i with $e_{\{i\}}$ and $e_\emptyset$ with the scalar $1 \in \mathbb{R}$.) It comes equipped with an associative bilinear $\wedge$-product which is completely determined by

$$\begin{aligned} e_i \wedge e_j &= -e_j \wedge e_i, && i, j \in [n], \\ e_{v_1} \wedge \ldots \wedge e_{v_k} &= e_{\{v_1, \ldots, v_k\}}, && 1 \leq v_1 < \ldots < v_k \leq n. \end{aligned}$$

Let $(e_A^*)_{A \subset [n]}$ be the dual basis of $(e_A)_{A \subset [n]}$. We naturally identify $\bigwedge(V^*)$ and $(\bigwedge V)^*$ so that $e_{v_1}^* \wedge \ldots \wedge e_{v_k}^*$ corresponds to $e_{\{v_1, \ldots, v_k\}}^*$, $1 \leq v_1 < \ldots < v_k \leq n$.

Let $\mathbf{f} = (f_1, \ldots, f_n)$ be another basis of V; in the obvious way we define f_A, f_A^* for $A \subset [n]$, etc. By $M = (\alpha_{ij})_{i,j \in [n]}$ we denote the $n \times n$-matrix satisfying $\mathbf{f}^* = M\mathbf{e}^*$, that is,

$$f_i^* = \alpha_{i1} e_1^* + \ldots + \alpha_{in} e_n^*, \qquad i \in [n].$$

Assume that $\mathbf{f}$ is in the *generic position* with respect to $\mathbf{e}$, that is, the entries of M are n^2 transcendentals algebraically independent over the rationals. An alternative definition is to assume that the entries are n^2 independent variables; any equation we will consider can be reduced to the form $P = 0$ for some polynomial P in the α's with integer coefficients and we agree that the statement is true if and only if P is the zero polynomial.

Let $\bigwedge^i V$ be the subspace of $\bigwedge V$ spanned by $(e_A)_{A\in[n]^{(i)}}$. We denote

$$\langle g^*, h\rangle = g^*(h), \quad g^* \in \bigwedge V^*,\ h \in \bigwedge V.$$

For $g^* \in \bigwedge V^*$, $h \in \bigwedge V$, the *left interior product* $g^* \llcorner h \in \bigwedge V$ is defined by

$$\langle u^*, g^* \llcorner h\rangle = \langle u^* \wedge g^*, h\rangle, \quad \text{for all } u^* \in \bigwedge V^*.$$

Thus, if $g^* \in \bigwedge^d V^*$ and $h \in \bigwedge^{d+l} V$ then $g^* \llcorner h \in \bigwedge^l V$, $d, l \geq 0$. One can easily check that $\llcorner$ is a bilinear function, such that $u^* \llcorner (g^* \llcorner h) = (u^* \wedge g^*) \llcorner h$ and, for the basis vectors, we have:

$$e_A^* \llcorner e_B = \begin{cases} \pm e_{B\setminus A}, & \text{if } A \subset B, \\ 0, & \text{if } A \not\subset B. \end{cases}$$

(The actual signs of ± 1-coefficients do not interest us at all.) Note that by the generality of $\mathbf{f}$ we have $\langle f_F^*, e_E\rangle \neq 0$ for any $E, F \in [n]^{(r)}$. Moreover, for any $|E| = r$, $f \in \bigwedge^i V$ and $g \in \bigwedge^{r-i} V$, we have

$$\langle e_E^*, f \wedge g\rangle = \sum_{A\in E^{(i)}} \sigma_{A,E} \langle e_A^*, f\rangle \cdot \langle e_{E-A}^*, g\rangle, \tag{52}$$

where $\sigma_{A,E} = \pm 1$ depending on A and E.

For $h \in \bigwedge V$, its *support* is defined by

$$\operatorname{supp}(h) = \{A \subset [n] : e_A^*(h) \neq 0\}. \tag{53}$$

That is, to find $\operatorname{supp}(h)$, write $h = \sum_{A\subset[n]} c_A e_A$ and take those $A \subset [n]$ for which the corresponding coefficient is non-zero. If we take the support in the basis $\mathbf{f}$ we emphasize this by adding a subscript:

$$\operatorname{supp}_{\mathbf{f}}(h) = \{A \subset [n] : f_A^*(h) \neq 0\};$$

while the supp alone always means the support relative to $\mathbf{e}$ as defined by (53).

Note that the cancellation $(g^* \wedge e_A^*) \llcorner (h \wedge e_A) = g^* \llcorner h$ (which is not generally correct) can be applied if, for example, each $B \in \operatorname{supp}(h)$ is disjoint from A. We will use identities like this a few times without detailed explanations. (The best way to verify them is to check them for the basis vectors.)

8.1.2 Definitions

Let us describe how to construct the *gross matroid* $\mathcal{G}_G$ of an r-graph G of order n.

Identify the vertices of G with the basis $\mathbf{e} = \{e_1, \ldots, e_n\}$ of V. Let $Z \subset \bigwedge^r V$ be defined by the following linear relations:

$$Z = \{h \in \textstyle\bigwedge^r V : f_E^* \llcorner h = 0 \text{ for all } E \in E(G)\}. \qquad (54)$$

As $(f_E^*)_{E\in[n]^{(r)}}$ forms a basis for $\bigwedge^r V^*$, we conclude that the $e(G)$ relations defining Z are linearly independent so $\dim Z = \binom{n}{r} - e(G)$ and, in fact, Z is spanned by $\{f_E : E \in E(\overline{G})\}$.

We define the *gross matroid* $\mathcal{G}_G$ on $[n]^{(r)}$ so that an r-graph F on $[n]$ is dependent if, for some coefficients c_E (not all zero), we have $\sum_{E\in E(F)} c_E e_E \in Z$. To verify this condition we have to find a non-zero solution $(c_E)_{E\in E(F)}$ of the following system of $e(G)$ linear equations:

$$\sum_{E\in E(F)} c_E f_D^* \llcorner e_E = 0, \qquad D \in E(G). \qquad (55)$$

By $M(G, F)$ we denote the $e(G) \times e(F)$-matrix corresponding to (55). The columns of $M(G, [n]^{(r)})$ provide a representation of $\mathcal{G}_G$. Note that the matroid $\mathcal{G}_G$ does not depend on the choice of generic $\mathbf{f}$. Also, $\mathcal{G}_G$ is a *symmetric* matroid, that is, for any permutation $\sigma : [n] \to [n]$, $A \subset [n]^{(r)}$ is $\mathcal{G}_G$-independent if and only if $\sigma'(A)$ is, where σ' is the induced action on $[n]^{(r)}$. Therefore, we can apply the notion of $\mathcal{G}_G$-dependence to an r-graph F with any vertex set. (If $v(F) > v(G)$, we add isolated vertices to G.)

This construction is not new; Kalai [Kal90] used it to construct symmetric matroids with a given growth polynomial. Also, in the partial case $G = P_{k,n-k}$, the matroid $\mathcal{G}_G$ is exactly Kalai's [Kal85] *k-hyperconnectivity matroid* on $[n]^{(2)}$ which was used to compute the w-sat-function for complete graphs. These two papers by Kalai were the starting points of our research on gross matroids.

Clearly, the rank of $\mathcal{G}_G$ is $\mathrm{codim}(Z) = e(G)$. It is easy to show that G is a base of $\mathcal{G}_G$. Indeed, the determinant of $M(G, G)$ is a polynomial in the α's which assumes value 1 when M (and then $M(G, G)$) is the identity matrix. Therefore, the determinant is non-zero for a generic M and the columns of $M(G, G)$ are independent, which proves the claim.

An r-graph F is a $\mathcal{G}_G$-chain if every $E \in E(F)$ is dependent on $E(F) \setminus \{E\}$ in $\mathcal{G}_G$, that is, for some $h \in Z$ and real c's, we have

$$e_E = h + \sum_{D \in E(F) \setminus \{E\}} c_D e_D. \tag{56}$$

This is easily seen to be equivalent to the existence of $h \in Z$ with $\mathrm{supp}(h) = E(F)$. To verify the last condition we have to find a solution $(c_E)_{E \in E(F)}$ with all entries non-zero of the system (55).

8.2 Count Matroids

Here we present the definition of a *count matroid* and establish some its properties. We generalize naturally the original definition of White and Whiteley [WW84] to obtain a considerably wider family of matroids for which we preserve the same name. For example, our count matroids admit many polynomials in n as the rank function while the original definition is confined to linear functions only. An advantage of count matroids is that they are defined in purely combinatorial terms and it is usually easy to identify their independent sets and circuits.

Count matroids are helpful in computing the w-sat-function, as is described in Section 9. We hope that they will have many other interesting applications; one is presented by Whiteley [Whi89].

8.2.1 Definitions

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

$$\rho(A \cup B) + \rho(A \cap B) \le \rho(A) + \rho(B), \quad A, B \in X^{(<\infty)}. \tag{57}$$

Given $\rho : X^{(<\infty)} \to \mathbb{R}$, 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 [ER66] observed the following result. (The proof is easy and can be found, for example, in Oxley [Oxl92, Proposition 12.1.1].)

Lemma 29 *For any integral increasing submodular function $\rho : X^{(<\infty)} \to \mathbb{R}$, the family of ρ-balanced sets satisfies the circuit axioms and therefore defines a matroid on X.* ▮

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 [WW84], see also [Whi96], introduced a family of *count matroids* on $[n]^{(r)}$ by defining

$$\rho(H) = a_1 | \cup_{E \in H} E| + a_0, \quad H \subset [n]^{(r)},$$

for some fixed a_1 and a_0.

We have found it possible to generalize 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

$$L(H) = a_0 + \sum_{i=1}^{r-1} a_i p_i(H), \quad H \subset [n]^{(r)}, \tag{58}$$

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

Let us see when the function L satisfies the above properties for $X = \mathbb{N}^{(r)}$. It is easy to see that L is integral if and only if all coefficients are integers. Submodular and increasing linear functions are characterized by the following two lemmas which are of independent interest.

Lemma 30 *A linear function $L : X^{(<\infty)} \to \mathbb{R}$ is increasing if and only if*

$$\sum_{j=i}^{r-1} a_j \binom{r}{j} \geq 0, \qquad i \in [r-1]. \tag{59}$$

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 (59).

On the other hand, suppose that L satisfies (59). 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,

$$0 \leq d_1 \leq \ldots \leq d_{r-1} \leq 1. \tag{60}$$

It is easy to check that

$$L(H \cup \{E\}) - L(H) = \sum_{i=1}^{r-1} a_i d_i \binom{r}{i}. \tag{61}$$

Consider the problem of minimizing (61) given only the constraints (60). 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 (61) is non-negative by (59), so L is increasing. ∎

Lemma 31 *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 (57) 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 $a_i \geq 0$. ∎

Thus we restrict our attention to integer coefficients satisfying

$$a_i \geq 0,\ i \in [r-1], \quad \text{and} \quad \sum_{j=0}^{r-1} a_j \binom{r}{j} \geq 1, \tag{62}$$

in which case, by Lemma 29, 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 (62) 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 σ' denotes the induced action on $[n]^{(r)}$. Clearly, the nested sequence $(\mathcal{N}_L^n)_{n\geq r}$ is compatible so we do not usually 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 [PW70]; this applies to all count matroids with $a_0 \geq 0$. We do not know if $\mathcal{N}_L$ is representable for $a_0 < 0$.

8.2.2 Rank

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

Theorem 32 *Let L satisfy* (62). *Then* $R(\mathcal{N}_L^n) = \min\left(\binom{n}{r}, L([n]^{(r)})\right)$.

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 $\overline{G} = \{E_1, \ldots, E_s\}$ such that

$$F_{[j-1]} \cap F_j \neq \emptyset, \quad j \in [2, s], \tag{63}$$

where F_i denotes the (unique and, by (62), 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 show the claim choose arbitrary $E_1 \in \overline{G}$ and, inductively, take for E_j any available edge satisfying (63). Suppose, on the contrary, that we are stuck after having chosen $E_1, \ldots, E_{j-1}$, 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 $(C \cap H_1) \cup G_2$ wo not decrease if we remove E. Therefore, there is a circuit $C' \ni E$ such that $C' \cap H_1 \subset C$ and $C' \cap H_2 \subset G_2$. 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 $p_i(C)$ will not increase by the Kruskal-Katona Theorem [Kru63, Kat66], 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 (62) 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:

$$L(F_{[j]} + E_{[j]}) \quad \leq \quad L(F_{[j-1]} + E_{[j-1]}) + L(F_j + E_j) - L(F_{[j-1]} \cap F_j)$$

$$\leq \ e(F_{[j-1]}) + e(F_j) - e(F_{[j-1]} \cap F_j) \ = \ e(F_{[j]}).$$

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, $L([n]^{(r)}) = L(G) = e(G) = R(\mathcal{N}_L^n)$. ∎

Remark. Kalai [Kal90] 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 and characterized all possible polynomials. Unfortunately, these are not confined to $L([n]^{(r)})$ with some L satisfying (62). For example, the k-hyperconnectivity matroid on $\mathbb{N}^{(2)}$ introduced by Kalai [Kal85] 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 [Kal90] are constructed by means of multilinear algebra.)

9 Proof Techniques

Here we present a few different methods for proving lower bounds on w-sat$(n, \mathcal{F})$. Of these, C-proofs and G-proofs can be viewed as sufficient criteria for $G_n \in$ w-SAT(n, F) to be of the minimal size. Our approach is based on gross and count matroids which are defined in Section 8.

The links with matroid theory are not surprising insofar as the definition of weak saturation suggests some kind of dependence; loosely speaking, an F-proper addition of edges corresponds to closure and the notion of a minimum weakly saturated graph resembles that of a base.

The following observation, due to Kalai [Kal85], is crucial to our work. Suppose that we have a matroid $\mathcal{M}$ on $[n]^{(r)}$ and an r-graph F which is an $\mathcal{M}$-*chain*, that is, for every embedding $V(F) \subset [n]$, every edge $E \in E(F) \subset [n]^{(r)}$ is dependent on $E(F) \setminus \{E\}$. Then we claim that the size of any weakly F-saturated graph G on $[n]$ is at least $R_{\mathcal{M}}([n]^{(r)})$, the rank of $\mathcal{M}$. Indeed, let $E_1, \ldots, E_k$ be an F-proper ordering of $E(\overline{G})$. By the definition, for every $i \in [k]$, there is an F-subgraph of $G_i = G + E_1 \ldots + E_i$ containing E_i. Thus, E_i lies in the

$\mathcal{M}$-closure of G_{i-1}, which inductively implies that G spans $[n]^{(r)}$ in $\mathcal{M}$, and the claim follows.

Clearly, the above argument can be applied to a family $\mathcal{F}$ of forbidden r-graphs.

Lemma 33 (Kalai) *We have*

$$\text{w-sat}(n,\mathcal{F}) \geq l, \tag{64}$$

if we can find a matroid $\mathcal{M}$ on $[n]^{(r)}$ such that every $F \in \mathcal{F}$ is an $\mathcal{M}$-chain and $R_{\mathcal{M}}([n]^{(r)}) \geq l$. ∎

In this case we say that we can *m-prove* the inequality (64). If, furthermore, $\mathcal{M}$ is a count matroid, a gross matroid, or a representable matroid, then (64) is said to be *c-proved*, *g-proved*, or *r-proved* correspondingly. Of course, if there exists $G \in \text{w-SAT}(n,\mathcal{F})$ with $e(G) = l$, then G is extremal. In this case we say that $\mathcal{F}$ admits an *m-proof* for n. In the obvious way we define a *c-proof*, a *g-proof*, and an *r-proof*.

Given a matroid $\mathcal{M}$ on $[n]^{(r)}$ and an r-graph F, let

$$D_{\mathcal{M}}(F) = \min_{F \subset [n]} \left(e(F) - R_{\mathcal{M}}(E(F))\right),$$

that is, for every embedding $F \subset [n]$, we compute how many F-edges can be removed without decreasing the $\mathcal{M}$-rank of $E(F)$ and take the minimum over all embeddings $F \subset [n]$. For a family $\mathcal{F}$ of r-graphs, we define

$$D_{\mathcal{M}}(\mathcal{F}) = \min\{D_{\mathcal{M}}(F) : F \in \mathcal{F}\}. \tag{65}$$

The following refinement of Lemma 33 is also useful.

Lemma 34 *Suppose that, for some family $\mathcal{F}$ of r-graphs and a matroid $\mathcal{M}$ on $[n]^{(r)}$, every $F \in \mathcal{F}$ is an $\mathcal{M}$-chain. Then,*

$$\text{w-sat}(n,\mathcal{F}) \geq R_{\mathcal{M}}([n]^{(r)}) + D_{\mathcal{M}}(\mathcal{F}) - 1. \tag{66}$$

Proof. As in Lemma 33, we conclude that $E(G)$ spans $[n]^{(r)}$ in $\mathcal{M}$ for any weakly $\mathcal{F}$-saturated graph G on $[n]$. Consider the first edge E added to G. It creates some forbidden $F \subset [n]$; clearly, $E(F) \setminus \{E\} \subset E(G)$. Therefore, there are $D_{\mathcal{M}}(F)-1$ edges in G which are dependent on the remaining edges and the lemma follows. ∎

We say that (66) is *m'-proved.* If $\mathcal{M}$ is a count, gross, or representable matroid, then we respectively *c'-prove*, *g'-prove*, or *r'-prove* (66). If the lower bound in (66) is sharp, then we obtain an *m'-proof.* In the obvious way we define a *c'-proof*, a *g'-proof*, and an *r'-proof.*

The characteristic feature of w-sat-type problems is that, given $\mathcal{F}$, it is usually fairly easy to come up with a correct example of a weakly $\mathcal{F}$-saturated graph G (as a rule, there are many different extremal graphs) and the harder part is to prove that G is minimal. So, a typical problem is, given $G \in \text{w-SAT}(n, \mathcal{F})$, to verify whether $e(G) = \text{w-sat}(n, \mathcal{F})$, that is, we want to have some sufficient and/or necessary conditions that a weakly $\mathcal{F}$-saturated graph G has the minimal number of edges. Even if there exists an m-proof, it is not obvious at all how to search for a suitable matroid.

However, the gross matroid of G seems a good candidate for $\mathcal{M}$. If each element of $\mathcal{F}$ is a $\mathcal{G}_G$-chain, then we immediately conclude that G is extremal. In this case say that the pair $(\mathcal{F}, G)$ admits a *G-proof.* Hence, the G-proof can be viewed as a sufficient criterion for $G \in \text{w-SAT}(n, F)$ to be of the minimal size.

As gross matroids are representable, we have the following 'hierarchy' of proofs (and other implications):

$$G\text{-proof} \;\Rightarrow\; g\text{-proof} \;\Rightarrow\; \text{r-proof} \;\Rightarrow\; \text{m-proof}.$$

Unfortunately, gross matroids are not, in general, very easy to handle; it takes some efforts to identify their chains. Also, there are many examples of eligible pairs which do not accept a G-proof. For example, minimum weakly K_3^2-saturated graphs are trees, of which only stars produce a G-proof. Besides, $G/g/g'$-proofs do not provide an immediate characterization of minimum weakly saturated graphs, as usually there seems to be no easy combinatorial description of the set of bases of a gross matroid.

However, many new results are proved here using gross matroids. Let us prove one trivial lemma which, when combined with the results of Sections 11 and 12, has non-trivial consequences.

Lemma 35 *Let $K = lK_r^r$ be the union of l disjoint r-edges. Then $\mathcal{G}_K$ is the* uniform matroid *of rank l, that is, an r-graph F is independent in $\mathcal{G}_K$ if and only if $e(F) \leq l$.*

In particular, for any family $\mathcal{F}$ of r-graphs and for any n with $\binom{n}{r} \geq l$, we can g-prove that w-sat$(n, \mathcal{F}) \geq l$*, where $l = \min\{e(F) : F \in \mathcal{F}\} - 1$.*

Proof. Let us show, by induction on l, that any r-graph H of size l is $\mathcal{G}_K$-independent. We may assume that $E = [r]$ is an edge in both these graphs. One can see that

$$\det(M(K, H)) = \pm\alpha_{11} \dots \alpha_{rr} \det(M(K', H')) + (\text{other terms}),$$

where H' and K' are obtained respectively from H and K by removing E and none of the 'other terms' contains $\alpha_{11} \dots \alpha_{rr}$ as a factor. By induction, we conclude that $\det(M(K, H)) \neq 0$, and the claim follows as the rank of $\mathcal{G}_K$ is $e(K) = l$. ∎

Count matroids can be applied to w-sat-type problems in the following, slightly different, way. Suppose that, for a range of values of n, we have $G_n \in$ w-SAT$(n, \mathcal{F})$ (conjectured to be extremal) such that $e(G_n)$ is a polynomial in n. Then we try to write explicitly the (unique, if it exists) count matroid $\mathcal{N}$ such that $R_{\mathcal{N}}([n]^{(r)}) = e(G_n)$ and check whether each $F \in \mathcal{F}$ is an $\mathcal{N}$-chain. If we succeed, then G_n is indeed extremal and we have a *C-proof.*

This approach is usually less successful than the one via gross matroids. Its weaknesses are that we must have a guess for a number of values of n and that not many polynomials are the growth polynomials of a count matroid. But still there are a few natural problems for which, of the above approaches, only count matroids produce results, e.g. for some uniform families, see Subsection 10.2.

10 Specific Classes

Here we obtain various results for certain particular forbidden families.

10.1 Pyramids

Here we compute the w-sat-function for *pyramids*, which includes a few interesting results as partial cases: for example, this proves formula (49) conjectured by Tuza [Tuz88, Conjecture 7].

Let t be fixed. Suppose we are given a sequence $\mathbf{r} = (r_1, \ldots, r_t)$ of non-negative integers and a sequence of disjoint sets $S_1, \ldots, S_t$ of sizes $\mathbf{s} = (s_1, \ldots, s_t)$ such that $r_{[i]} \le s_{[i]}$, $i \in [t]$. (Dealing with sequences, we use such shortcuts as $r_I = \sum_{i \in I} r_i$ and $S_I = \cup_{i \in I} S_i$, $I \subset [t]$; we also assume $r_0 = 0$, $S_0 = \emptyset$.)

The *pyramid* $P = P(\mathbf{s}; \mathbf{r})$ is the r-graph, $r = r_{[t]}$, on $S = S_{[t]}$ such that $E \in S^{(r)}$ is an edge of P if and only if, for every $i \in [t]$, we have $|E \cap S_{[i]}| \ge r_{[i]}$. Of course, this condition is vacuously true for $i = t$.

For example, for $t = 1$ we have complete graphs; $P(s_1, s_2; r_1, r_2)$ consists of those $(r_1 + r_2)$-subsets of $S_1 \cup S_2$ which intersect S_1 in at least r_1 vertices. As a warning, we emphasize that pyramids are usual (not layered) r-graphs.

Without loss of generality we may assume that $s_i \ge r_i$, $i \in [t]$. If some r_i exceeds s_i then, letting $\mathbf{r}' = \mathbf{r}$ except $r'_i = s_i$ and $r'_{i-1} = r_{i-1} + r_i - s_i$ (note that $i \ge 2$ as $r_1 \le s_1$), we obtain the same pyramid $P' = P$. Indeed, $r_{[j]}$'s do not change except $r'_{[i-1]} = r_{[i-1]} + r_i - s_i$, so, trivially, $P' \subset P$. On the other hand, $E \in E(P)$ implies that

$$|E \cap S_{[i-1]}| \ge |E \cap S_{[i]}| - s_i \ge r_{[i]} - s_i = r'_{[i-1]},$$

and $E \in E(P')$. Iterating the step as long as possible, we prove the claim.

Likewise we can get rid of $r_i = 0$ by merging S_i and S_{i+1} together (or removing S_t if $i = t$).

Here we calculate w-sat(n, P) by showing that pyramids admit a G-proof. Note that we obtain the exact answer for *all feasible* values of the parameters n, $\mathbf{r}$ and $\mathbf{s}$. This result appears in [Pik99a].

Let us, for any $n \ge s = s_{[t]}$, provide a construction of $G \in$ w-SAT(n, P). Partition $[n] = A_1 \cup \ldots \cup A_{t+1}$ so that $a_i = |A_i| = s_i + r_{i-1} - r_i$, $i \in [t]$; thus

$$a_{t+1} = |A_{t+1}| = n - \sum_{i=1}^{t} (s_i + r_{i-1} - r_i) = n - s + r_t.$$

We also assume that our partition is *consecutive*, that is, in $[n]$, any element of A_i comes before any element of A_j whenever $i < j$.

Let $E \in [n]^{(r)}$ be an edge of G if and only if, for some $i \in [t]$, we have $|E \cap A_{[i]}| > r_{[i-1]}$. Equivalently, the complement of G is isomorphic to

$$P(a_{t+1}, \ldots, a_1; r_t, \ldots, r_1, 0),$$

so, for example, any r-tuple intersecting A_1 is in $E(G)$.

Lemma 36 $G \in$ w-SAT(n, P).

Proof. Order the missing edges in any way so that the sequences

$$\left(|A_{[1]} \cap E|, \ldots, |A_{[t+1]} \cap E|\right), \quad E \in E(\overline{G}),$$

are non-increasing in the lexicographic order. (Thus, we start with $(0, r_1, \ldots, r_t)$ and end with $(0, \ldots, 0, r)$.) Let us show that this ordering is P-proper. Consider the moment when we add some edge $E \in E(\overline{G})$. Let $E_i = E \cap A_{i+1}$, $i \in [t]$. Also, let $E = R_1 \cup \ldots \cup R_t$ and $[n] \setminus E = T_1 \cup \ldots \cup T_{t+1}$ be the consecutive partitions with $|R_i| = r_i$ and $|T_i| = s_i - r_i$, $i \in [t]$.

Let us show that $E_{[i]} \subset R_{[i]}$ and $T_{[i]} \subset A_{[i]} \setminus E_{[i-1]}$, $i \in [t]$. As all partitions in question are consecutive, it is enough to verify the sizes. By the definition of G, we have $|E_{[i]}| = |E \cap A_{[i+1]}| \le r_{[i]}$. Also,

$$|A_{[i]} \setminus E_{[i-1]}| \ge |A_{[i]}| - r_{[i-1]} = \sum_{j=1}^{i} (s_j + r_{j-1} - r_j) - r_{[i-1]} = |T_{[i]}|,$$

and the claim follows.

Let $S_i = T_i \cup R_i$, $i \in [t]$. We claim that E creates a forbidden subgraph P on the set $S = S_{[t]}$. For every $i \in [t]$, we have $|E \cap S_i| = |R_i| = r_i$, so $E \in E(P)$.

Suppose, on the contrary, that there exists $D \in E(P)$ coming after E. Let us show by induction on i that, for every $i \in [0, t]$, we have

$$D \cap S_{[i]} = E \cap S_{[i]} \quad \text{and} \quad D \cap A_{[i+1]} = E \cap A_{[i+1]}, \tag{67}$$

which would be a contradiction to the assumption $D \neq E$. As $D, E \in E(\overline{G})$ are disjoint from A_1, the claim is true for $i = 0$. Let $i \in [t]$. As $T_{[i]} \subset A_{[i]}$, we conclude, by the inductive assumption, that $D \cap T_{[i]} = E \cap T_{[i]} = \emptyset$. As $S_{[i]} = T_{[i]} \cup R_{[i]}$, we have $D \cap S_{[i]} \subset R_{[i]}$. On the other hand, $D \in E(P)$ so $|D \cap S_{[i]}| \ge r_{[i]}$, which implies

$$D \cap S_{[i]} = R_{[i]} = E \cap S_{[i]},$$

and the first part of (67) is proved. Now,

$$D \cap A_{[i+1]} \supset R_{[i]} \cap A_{[i+1]} \supset E_{[i]} \cap A_{[i+1]}.$$

By induction, $D \cap A_{[i]} = E \cap A_{[i]}$ and, as D was added later than E, we must have $|D \cap A_{[i+1]}| \leq |E \cap A_{[i+1]}|$, which proves (67) completely. ∎

Theorem 37 *The pair (P, G) admits a G-proof.*

Proof. We have to show that P is a $\mathcal{G}_G$-chain. Let us consider

$$h = h_1 \wedge \ldots \wedge h_t, \quad \text{where} \;\; h_i = f^*_{A_{[i]}} \llcorner e_{S_{[i]}} \in \bigwedge^{r_i} V, \;\; i \in [t],$$

where, as usual, $\mathbf{f}^*$ is a generic V^*-basis relative a V-basis $\mathbf{e}$. Each $E \in \operatorname{supp}(h)$ is of the form $E_1 \cup \ldots \cup E_t$, for some $E_i \in \operatorname{supp}(h_i)$, $i \in [t]$. Clearly, $|E_i| = r_i$ and $E_i \subset S_{[i]}$. Therefore, $|E \cap S_{[i]}| \geq |E_{[i]}| = r_{[i]}$, so $\operatorname{supp}(h) \subset E(P)$. Similarly, $\operatorname{supp}_{\mathbf{f}}(h_i)$ lives within $A_{[i+1,t+1]}$, $i \in [t]$, which implies that $\operatorname{supp}_{\mathbf{f}}(h) \subset E(\overline{G})$.

So, to prove the theorem, it is enough to show that for any $E \in E(P)$ we have $P_E = \langle e^*_E, h \rangle \neq 0$. To do so, we can assume that S is an initial segment in $[n]$ and every element of S_i comes before every element of S_j whenever $i < j$. Furthermore, we may assume that $E_i = E \cap S_i$ is a final segment of S_i. Note that $A_{[i]} \subset S_{[i]} \subset A_{[i+1]}$ and $R_i = S_{[i]} \setminus A_{[i]}$ consists of the last r_i elements of S_i, $i \in [t]$. Clearly, $|E| = |R|$, where $R = R_{[t]}$, so let $g : E \setminus R \to R \setminus E$ be the order-preserving bijection.

As P_E is a polynomial in the α's, to show that $P_E \neq 0$, it is enough to demonstrate a particular example of the α's (or $\mathbf{f}^*$) such that $P_E \neq 0$. Define

$$f^*_x = \begin{cases} e^*_x + e^*_{g(x)}, & x \in E \setminus R, \\ e^*_x, & \text{otherwise.} \end{cases} \tag{68}$$

Let $i \in [t]$. To compute h_i, we expand $f^*_{A_{[i]}}$ in the $\mathbf{e}^*$-basis by (68). Denote $W_i = A_{[i]} \setminus (E \setminus R)$ and

$$\begin{aligned} X_i &= \{x \in A_{[i]} \setminus W_i : g(x) \in A_{[i]}\}, \\ Y_i &= \{x \in A_{[i]} \setminus W_i : g(x) \notin S_{[i]}\}, \\ Z_i &= \{x \in A_{[i]} \setminus W_i : g(x) \in S_{[i]} \setminus A_{[i]}\}. \end{aligned}$$

As $A_{[i]} \subset S_{[i]}$ we have a partition $A_{[i]} = W_i \cup X_i \cup Y_i \cup Z_i$. As $f^*_x = e^*_x$ for $x \in W_i$,

$$f^*_{A_{[i]}} = \pm f^*_{W_i} \wedge f^*_{X_i} \wedge f^*_{Y_i} \wedge f^*_{Z_i} = \pm e^*_{W_i} \wedge f^*_{X_i} \wedge f^*_{Y_i} \wedge f^*_{Z_i}.$$

Take some $x \in X_i$; then $g(x) \in W_i$. Now, for some $u^* \in \bigwedge V^*$, we have the following representation

$$f^*_{A_{[i]}} = f^*_x \wedge f^*_{g(x)} \wedge u^* = (e^*_x + e^*_{g(x)}) \wedge e^*_{g(x)} \wedge u^* = e^*_x \wedge e^*_{g(x)} \wedge u^*,$$

which implies that $f^*_{A_{[i]}} = \pm e^*_{W_i} \wedge e^*_{X_i} \wedge f^*_{Y_i} \wedge f^*_{Z_i}$.

Next, consider some $x \in Y_i$; then $g(x) \notin S_{[i]}$. For some $u^* \in \bigwedge V^*$, we have

$$f^*_{A_{[i]}} \llcorner e_{S_{[i]}} = (u^* \wedge f^*_x) \llcorner e_{S_{[i]}} = \Big(u^* \wedge (e^*_x + e^*_{g(x)})\Big) \llcorner e_{S_{[i]}} = (u^* \wedge e^*_x) \llcorner e_{S_{[i]}},$$

that is, we can replace f^*_x by e^*_x without affecting h_i. Also $g(Z_i) \cap A_{[i]} = \emptyset$ and $S_{[i]} \setminus A_{[i]} = R_i$, so

$$h_i = \pm(e^*_{W_i} \wedge e^*_{X_i} \wedge e^*_{Y_i} \wedge f^*_{Z_i}) \llcorner e_{S_{[i]}} = \pm f^*_{Z_i} \llcorner e_{Z_i \cup R_i}.$$

For $i \in [t]$, we have $|E_{[i-1]}| \geq |R_{[i-1]}|$ and one of E_i and R_i is a subset of the other, so, for each $x \in E_i \setminus R$, $g(x)$ lies in $R_j = S_{[j]} \setminus A_{[j]}$ and $x \in Z_j$, for some $j \in [i+1, t]$. Therefore, $Z_{[t]} = E \setminus R$.

When we compute $P_E = \pm\langle e^*_E, \wedge_{i\in[t]}(f^*_{Z_i} \llcorner e_{Z_i \cup R_i})\rangle$ by expanding further each h_i in the $\mathbf{e}$-basis, we obtain h as a sum of terms each of the form e_D, for some $D \in [n]^{(r)}$. By definition, $\langle e^*_E, e_D\rangle = 0$ unless $E = D$. Consider some $x \in Z_i \subset E$. As $x \notin R$ and $Z_1, \ldots, Z_t$ are disjoint, no element of $\mathrm{supp}(h_j)$ can contain x unless $j = i$. Computing h_i, we have for some u^*

$$h_i = (u^* \wedge f^*_x) \llcorner e_{Z_i \cup R_i} = (u^* \wedge e^*_{g(x)}) \llcorner e_{Z_i \cup R_i} + (u^* \wedge e^*_x) \llcorner e_{Z_i \cup R_i},$$

and no element in the $\mathbf{e}$-support of the second summand can contain x. Thus we can harmlessly replace f^*_x by $e^*_{g(x)}$. (Clearly, this does not affect h_j for $j \neq i$.) Now, since $g(Z_i) \subset S_{[i]} \setminus A_{[i]} = R_i$,

$$\begin{aligned} P_E &= \pm\langle e^*_E, \wedge_{i\in[t]}(e^*_{g(Z_i)} \llcorner e_{Z_i \cup R_i})\rangle \\ &= \pm\langle e^*_E, e_{Z_{[t]} \cup R_{[t]} \setminus g(Z_{[t]})}\rangle \;=\; \pm\langle e^*_E, e_E\rangle \;=\; \pm 1. \end{aligned}$$

Thus P_E is non-zero and the theorem follows. ∎

Corollary 38 *Suppose that we are given two sequences* $\mathbf{s} = (s_1, \ldots, s_t)$ *and* $\mathbf{r} = (r_1, \ldots, r_t)$ *of integers such that* $s_i \geq r_i \geq 1$, $i \in [t]$. *Then, for* $n \geq s_{[t]}$,

$$\text{w-sat}(n, P(\mathbf{s};\mathbf{r})) = \binom{n}{r_{[t]}} - \sum_{\mathbf{r}'} \binom{n - s_{[t]} + r_t}{r'_t} \prod_{i\in[t-1]} \binom{s_{i+1} + r_i - r_{i+1}}{r'_i},$$

where the summation is taken over all sequences of non-negative integers $\mathbf{r}' = (r'_1, \ldots, r'_t)$ *such that* $r'_{[t]} = r_{[t]}$ *and, for any* $i \in [t-1]$, $r'_{[i]} \le r_{[i]}$. ∎

Remark. To achieve equality in Corollary 38, the edges of a weakly P-saturated graph H must form a base in $\mathcal{G}_G$. As it is the case with G/g-proofs, there is no easy combinatorial interpretation of this condition.

Pyramids cover many interesting graphs as partial cases and Corollary 38 implies new results even for $r = 2$: we are able to compute the w-sat-function for $P_{s,t} = P(s,t;1,1)$, the disjoint union of K_s^2 and E_t^2 plus all edges between them. Namely, for $n \ge s+t$, $s \ge 1$, $t \ge 1$, we have

$$\text{w-sat}(n, P_{s,t}) = (s-1)n - \binom{s}{2} + \binom{t}{2}.$$

As $P(m;r) = K_m^r$, we can compute w-sat(n, K_m^r), formula (46) here.

Observe that $P(r-l+1, l; r-l+1, l-1)$ is the only member of $\mathcal{H}_r(r+1,l)$, which proves the formula (49) conjectured by Tuza [Tuz88, Conjecture 7].

Also, $S_m^r = P(1, m-1; 1, r-1)$. Therefore, Corollary 38 directly implies that

$$\text{w-sat}(n, S_m^r) = \binom{n}{r} - \binom{n-k}{r} - k\binom{n-k}{r-1}, \quad n \ge m > r \ge 2,$$

where $k = m-r+1$. A complete description of all minimum weakly S_m^r-saturated graphs is available only for S_m^2 when we can find a simple combinatorial proof which, fortunately, works for the following, wider, class of graphs.

A *delta system* D_{ml}^r contains l r-tuples so that the intersection of every two is equal to a fixed m-set called the *centre*. Thus, $v(D_{ml}^r) = m + l(r-m)$.

Theorem 39 *For any* $r > m \ge 1$ *and* $n > m + l(r-m)$, w-sat$(n, D_{ml}^r) = \binom{l}{2}$.

Proof. To construct $G \in$ w-SAT(n, D_{ml}^r), choose $A \in [n]^{(m-1)}$ and distinct vertices $y_1, \ldots, y_{l-1} \in [n] \setminus A$. For each $i \in [l-1]$, place into $E(G)$ any $l-i$ edges forming a $D_{m,l-i}^r$-graph centred at $A \cup \{y_i\}$ and disjoint from $\{y_1, \ldots, y_{i-1}\}$.

Let us show that $G \in$ w-SAT(n, D_{ml}^r). Repeat the following step for $i = 1, \ldots, l-1$. Suppose, we have already added to G all edges containing A and intersecting $\{y_1, \ldots, y_{i-1}\}$. Observe that by now we have a $D_{m,l-1}^r$-subgraph centred at $A \cup \{y_i\}$. It is not hard to check that we can properly add to G all edges containing $A \cup \{y_i\}$, cf. Theorem 52.

Finally, add, in any order, the remaining edges so that $|E \cap A|$ is non-increasing. Easy details are omitted.

Conversely, given $G \in$ w-SAT(n, D^r_{ml}), define inductively $A_1, \ldots, A_{l-1} \subset V(G)$ as follows. For $i = 1, \ldots, l-1$, consider the first edge added to G containing none of $A_1, \ldots, A_{i-1}$ as a subset. Let A_i be the centre of a created D^r_{ml}-subgraph F. For any $j \in [i-1]$, at most one edge of F can contain A_j because any two such edges overlap in $A_i \cup A_j$ which has size at least $m+1$. Therefore, at least $l - i$ edges of F belonged to the initial G. These edges contain A_i but none of $A_1, \ldots, A_{i-1}$. So, $e(G) \geq (l-1) + (l-2) + \ldots + 1 = \binom{l}{2}$. ∎

Remark. It is easy to read off the proof the characterization of all extremal graphs for $S^2_m = D^2_{1,m-1}$ (and for some other cases): all minimum weakly S^2_m-saturated graphs can be obtained in the following way. Choose $\{x_1, \ldots, x_{m-2}\} \in [n]^{(m-2)}$. For every $i \in [m-2]$, add any $m - i - 1$ edges through the vertex x_i not incident to $x_1, \ldots, x_{i-1}$.

10.2 Uniform Families

Fix $l, m, r \in \mathbb{N}$ with $1 \leq l \leq \binom{m}{r}$. 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, $G \in$ w-SAT$(n, \mathcal{H})$, $n \geq m$, if we can consecutively add the missing edges so that each creates a new subgraph with at most m vertices and at least l edges.

There are quite a few papers dealing with the Turán ex-function for uniform families; we refer the reader to Griggs, Simonovits and Thomas [GST98] for references and for new recent results.

The sat-type problems for uniform families were considered by Tuza [Tuz88], who made a conjecture about the value of w-sat$(n, \mathcal{H}_r(r+1, l))$ (formula (49) here), and by Erdős, Füredi and Tuza [EFT91] who settled the case $l = 3$ of Tuza's conjecture. Observe that we have essentially only one graph in $\mathcal{H}_r(r+1, l)$ which consists of all edges containing some fixed $(r-l+1)$-set. In our notation it is denoted by $P(r-l+1, l; r-l+1, l-1)$, and Corollary 38 implies formula (49).

However, the general case is still open.

Here we present, for all sets of parameters, a construction of a weakly $\mathcal{H}_r(m, l)$-saturated graph which we conjecture to be extremal. Our conjecture is in perfect accordance with the above results.

Clearly, our construction gives an upper bound. To establish some lower bounds, we use use gross and count matroids. This way we verify our conjecture for more sets of parameters. In certain cases, we characterize the sets of minimum weakly $\mathcal{H}$-saturated graphs. In particular, we answer a question by Erdős, Füredi and Tuza [EFT91] who asked for a characterization of the extremal graphs for $\mathcal{H}_r(r+1,3)$. These results appear in [Pik98].

10.2.1 Construction

Let $n \geq m$, $1 \leq l \leq \binom{m}{r}$ 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 we can take for G_n any member of $\mathcal{H}_r(m,l-1)$. If $n > m$, then choose inductively any $G_{n-1} = G(n-1,r,m,l)$ and $G' = G(n-1,r-1,m-1,l')$, where $l' = l - \binom{m-1}{r}$. (If $l \leq \binom{m-1}{r} + 1$ then we take the empty graph for G'.) Let G_n be the r-graph on $[n]$ defined by

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

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 $M \supset E_i$, $M \cup \{n\} \subset V(G)$ 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 40 *For any $n,r,m,l \in \mathbb{N}$ satisfying $m \leq n$ and $1 \leq l \leq \binom{m}{r}$, $G(n,r,m,l)$ is a minimum weakly $\mathcal{H}_r(m,l)$-saturated graph.*

Remark. Generally, not all extremal graphs are given by our construction, cf. Theorem 44.

Let us compute the size of G_n. Given $l \geq 2$, 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.

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

(We agree that $\binom{i}{0} = 1$, for *any* i.) For our purposes, we have to find a representation of the form $e(G_n) = \sum_{k=0}^{d} a_k \binom{n}{k}$. The 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. [GKP89, 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 $\sum_{i=0}^{t} (-1)^i \binom{j}{i} = (-1)^t \binom{j-1}{t}$, $t \geq 0$, we can find that $a_k = (-1)^{d-k} c \binom{m-k-1}{d-k} + (-1)^k s_k$, where

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

Therefore, in summary,

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

One can check that Conjecture 40 is compatible with (44), which is one more point supporting Conjecture 40.

10.2.2 Applications of Count Matroids

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

$$a_k = (-1)^{d-k} \left(c \binom{m-k-1}{d-k} - \binom{m-k-1}{r-k} \binom{r-k-1}{d-k-1} \right). \qquad (69)$$

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 w-sat$(n, \mathcal{H}_r(m, l)) = e(G_n)$, which establishes the validity of our conjecture in this case.

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.

Case 1: $\mathbf{d = 0}$. In this case the problem is trivial: it is easy to prove directly the following result (also observed by Erdős, Füredi and Tuza [EFT91]).

Lemma 41 *For $n \geq m \geq r \geq 1$ and $1 \leq l \leq \binom{m-1}{r} + 1$,*

$$\text{w-sat}(n, \mathcal{H}_r(m,l)) = l - 1.$$

All extremal graphs are can be obtained by adding $n - m$ isolated vertices to an $F \in \mathcal{H}_r(m, l-1)$. (Which is exactly what our construction says.) ∎

Case 2: $\mathbf{d = 1}$. Let $l = \binom{m-1}{r} + 1 + c$, $1 \leq c \leq \binom{m-2}{r-1}$. By (69), we let $a_1 = c$ and $a_0 = \binom{m-1}{r} - c(m-1)$, that is,

$$L(H) = cp_1(H) + \binom{m-1}{r} - c(m-1), \quad H \subset [n]^{(r)}.$$

The condition $1 \leq a_1 r + a_0$ implies that either $m = r+1$ (then $c \leq \binom{m-2}{r-1}$ must equal 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 32 as $L([m-1]^{(r)}) = \binom{m-1}{r}$. Hence F is a circuit and our conjecture is true.

Lemma 42 *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* $\text{w-sat}(n, \mathcal{H}_r(m,l)) = (l-1) + c(n-m)$. ∎

In some cases, we can characterize extremal graphs by providing a combinatorial proof.

Lemma 43 *In addition to the assumptions of Lemma 42, assume that* $m > r+1$ *and* $c < \frac{1}{m-1}\binom{m-1}{r}$. *Then any minimum* $G \in \text{w-SAT}(n, \mathcal{H}_r(m,l))$ *is given by our construction.*

Proof. Let $\overline{G} = \{E_1, \ldots, E_s\}$ be a proper ordering; suppose 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 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

$$e(H_i) - e(H_{i-1}) \geq l - \binom{m-k}{r} - 1 = c + \binom{m-1}{r} - \binom{m-k}{r}.$$

It is routine to check that the last expression is strictly greater than ck 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. ∎

As we mentioned, the value of $\text{w-sat}(n, \mathcal{H}_r(r+1, 3))$ was computed by Erdős, Füredi and Tuza [EFT91]. They asked if there is a characterization of the extremal graphs. Our Lemma 43 does not cover this case but we can provide a different proof of the lower bound which gives us the desired characterization. Some ideas from [EFT91] are used here but, of course, we have to be more delicate if we want to extract the cases of equality.

Theorem 44 *For* $\mathcal{H} = \mathcal{H}_r(r+1, 3)$ *we have*

$$\text{w-sat}(n, \mathcal{H}) = n - r + 1, \qquad n \geq r. \tag{70}$$

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 (70), we have to provide a combinatorial proof of the lower bound. Let $G \in$ 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 $E_1, \ldots, E_j$ be the edges of G. With this sequence we do, step by step and as long as possible, the following operation. If some 2 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 a sequence containing a single member (which then must be equal to $V(G)$). Suppose not. Let $Y_1, \ldots, Y_t$, $t \geq 2$, 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 (70), then, in each step of 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 intersection $A = V(H_1) \cap V(H_2)$ of size $r-1$, is weakly S-saturated. But this is easy: for $i = r-2, r-1, \ldots$, add the missing edges which intersect A in exactly i points. ∎

Remark. The construction of $G(n, r, r+1, 3)$ before Conjecture 40 does not cover all cases as is demonstrated, for example, by $r = 3$, $n = 6$ and

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

Case 3: $\mathbf{d = 2}$. 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 (69), 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 (62). 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

$$0 < c\binom{m-r-1}{2} + \left(r - \frac{(m-1)(r-1)}{r}\right)\binom{m-2}{r-1}. \tag{71}$$

Note that (71) 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 (62) if and only if either $m = r+2$ and $c = 1$ or $m \geq r+3$ and

$$\frac{((m-1)(r-1) - r^2)\binom{m-2}{r-1}}{r\binom{m-r-1}{2}} < c \leq \min\left(\frac{\binom{m-2}{r-1}}{m-2}, \binom{m-3}{r-2}\right) = \frac{\binom{m-2}{r-1}}{m-2}. \tag{72}$$

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 32, $[m-1]^{(r)}$ 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

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

If we let $L' = a_2p_1 + a_1$ then $L'([m-2]^{(r-1)}) = \binom{m-2}{r-1}$ and, by Theorem 32, $H' \subset [m-2]^{(r-1)}$ 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 45 *Assume that $r \geq 3$ and $l = \binom{m-1}{r} + \binom{m-2}{r-1} + c + 1$ are such that either $m = r+2$ and $c = 1$ or $m \geq r+3$ and c satisfies* (72)*. Then Conjecture 40 is true.* ▮

Remark. Unfortunately, we do not have any characterization of the extremal graphs in this case.

10.2.3 Applications of Gross Matroids

We establish some further results by applying gross matroids. Namely, we prove that our conjecture is asymptotically true for $d = r - 1$. Moreover, by applying the g'-method we settle completely the case $r = 2$.

First, we need one simple preliminary result.

Lemma 46 *Let G be an r-graph of order n and size at least $\binom{n}{r} - n + m$, where $n > m > r \geq 2$. Then any $E \in E(G)$ is contained in a complete subgraph of order m.*

Proof. Given $E \in E(G)$, remove from each missing edge one (arbitrary) vertex not belonging to E. We are left with at least m vertices spanning a complete subgraph which contains E. ▮

Remark. The above bound on $e(G)$ is sharp: if the complement of G consists of $n - m + 1$ edges containing some fixed $(r-1)$-set A then this set is covered only by $m - r$ G-edges of which none lies within K_m^r.

Theorem 47 *Let $l = \binom{m}{r} - k$ and $\mathcal{H} = \mathcal{H}_r(m, l)$. If $m > k + r$, then*

$$\text{w-sat}(n, \mathcal{H}) = (m - k - r)\binom{n}{r-1} + O(n^{r-2}). \tag{73}$$

Furthermore, if $r=2$, then we have a g'-proof that

$$\text{w-sat}(n, \mathcal{H}_2(m,l)) = (m-k-2)(n-m)+l-1, \quad n \geq m. \tag{74}$$

Proof. Implementing our construction, from the identity $\sum_{i=0}^{r}\binom{m-i-1}{r-i} = \binom{m}{r}$, we obtain that $d=r-1$ and $c=m-r-k$, which implies the upper bounds in (73) and (74).

On the other hand, in any $F \in \mathcal{H}$, any edge lies within a K_{m-k}^r-subgraph by Lemma 46. But by Theorem 37, K_{m-k}^r is a chain in $\mathcal{G}_P$, the gross matroid of $P = P(c, n-c; 1, r-1)$, so each $F \in \mathcal{H}$ is a $\mathcal{G}_P$-chain. By Lemma 33, $\text{w-sat}(n,\mathcal{H}) \geq R_{\mathcal{G}_P}([n]^{(r)}) = e(P)$, which g-proves the required lower bound in (73).

Finally, let us g'-prove the lower bound in (74) for $r=2$. Let $F \in \mathcal{H}$. As F has m vertices,

$$R_{\mathcal{G}_P}(F) \leq R_{\mathcal{G}_P}(K_m) \leq e(P(c, m-c; 1, 1)) = cm - \binom{c+1}{2}.$$

(The second inequality is true because $P(c, m-c; 1,1) \in \text{w-SAT}(m, K_{m-k}^2)$ and K_{m-k}^2 is a $\mathcal{G}_P$-chain.) Therefore some set of at least $p = l - cm + \binom{c+1}{2}$ edges of F lies in the $\mathcal{G}_P$-span of the remaining edges, that is, $D_{\mathcal{G}_P}(F) \geq p$. By Lemma 34,

$$\begin{aligned} \text{w-sat}(n,\mathcal{H}) &\geq R_{\mathcal{G}_P}(K_n) + D_{\mathcal{G}_P}(\mathcal{F}) - 1 \\ &\geq cn - \binom{c+1}{2} + p - 1 \;=\; c(n-m) + l - 1. \end{aligned}$$

The theorem is proved. ▮

Note that for $r=2$ we know $\text{w-sat}(n, \mathcal{H}_2(m,l))$ for any any feasible m and l: for $l \leq \binom{m-1}{2}$ we have a g-proof that it is $l-1$ (constant) by Lemma 35, while all other cases are covered by the g'-proof of Theorem 47.

Also note that, under the assumptions of Theorem 47 on l, the graph $G(n,r,m,l)$ constructed before Conjecture 40 is weakly F-saturated, where $E(F)$ consists of the first l elements of $[m]^{(r)}$ in the colex order. So, Theorem 47 remains valid if F is the only member of $\mathcal{H}$; this covers all possible cases for $r=2$ except the trivial case $l = \binom{m-1}{2}+1$.

10.3 Miscellaneous Graphs

Here we indicate a few easy results for some simple forbidden graphs such as cycles, disjoint edges, trees, etc. The proofs are easy but they often require a lengthy and boring verification that the specified graph is weakly saturated. We include them for the sake of completeness.

Cycles

Let C_l denote the cycle of length l. We know (see Section 2) that the determination of the exact value of $\mathrm{sat}(n, C_l)$ is a hard task. For the w-sat-function, on the contrary, the complete answer is available in all cases.

The following trivial observation will be used a few times, so we state it as a lemma.

Lemma 48 *Let $l \geq 4$ be even. Then any weakly C_l-saturated graph G contains an odd cycle.*

Proof. Indeed, otherwise G is a bipartite graph. Let E be the first added edge lying within one part. By the parity argument, any l-cycle through E must contain another edge lying within a part, which is a contradiction to the choice of E. ∎

Let us first consider the case when the forbidden cycle is Hamiltonian.

Theorem 49 *For any $n \geq 4$, $\text{w-sat}(n, C_n) = n$ and all extremal graphs are obtained from a Hamiltonian cycle by adding an edge which creates an odd cycle and then removing some other edge.*

Proof. Let G be a Hamiltonian cycle visiting the vertices $1, 2, \ldots, n \in [n]$ in this order, minus the edge $\{1, n\}$ but plus the edge $\{i, n\}$, for some even i. To prove that $G \in \text{w-SAT}(n, C_n)$ we have to show how to properly add the missing edges to G. First we add $\{1, n\}$ thus creating a Hamiltonian cycle through $1, 2, \ldots, n$.

We fix this cycle and define a *t-chord* as an edge connecting 2 vertices at a distance t if we go along the cycle. Thus, after the first step, G is made of all 1-chords and one i-chord. Next, we add all i-chords in the following order $\{m, i+m\}$, $m = 1, 2, \ldots, n-1$. (Of course, we do all arithmetic modulo n.) Every time we receive an extra cycle: for example, the chord $\{1, i+1\}$ creates the cycle via

$$n, i, i-1, i-2, \ldots, 2, 1, i+1, i+2, \ldots, n-2, n-1.$$

Having all chords of length 1 and $i \geq 4$, it is possible to add any $(i-2)$-chord. For example, the chord $\{2, i\}$ creates the following Hamiltonian cycle

$$1, i+1, i+2, 2, i, i-1, \ldots, 4, 3, i+3, i+4, \ldots, n-1, n.$$

Therefore, we can eventually have all 2-chords.

Finally, consequently for $m = 3, 4, 5, \ldots$, we add all missing m-chords in any order. This is legitimate; when we add, for example, the chord $\{1, m+1\}$ we have a Hamiltonian cycle via

$$1, 2, \ldots, m-1, n, n-1, \ldots, m+2, m, m+1$$

which uses only already present chords (of length 1, 2 and $m-1$). Therefore, $G \in \text{w-sat}(n, C_n)$.

On the other hand, suppose that $G \in \text{w-SAT}(n, C_n)$. The first edge added to G creates a C_n-subgraph F (that is, a Hamiltonian cycle), so there is a Hamiltonian path P_{n-1} in G. It is easy to see that $P_{n-1} \notin \text{w-SAT}(n, C_n)$ so there is at least one more edge E and $\text{w-sat}(n, C_n) \geq n$. Moreover, $F + E$ must contain at least one odd cycle by Lemma 48, which is precisely what our construction says. ∎

Let us consider odd and even cycles separately.

Theorem 50 *Let $l \geq 3$ be odd and let $n > l$. Then $\text{w-sat}(n, C_l) = n - 1$, all extremal graphs are trees of order n and diameter at least $l - 1$, and C_l admits a g-proof for n.*

Proof. Let G be any such tree. First we add any edge connecting two vertices at distance $l-1$; suppose the created l-cycle goes through the vertices $1, \ldots, l \in V(G)$ in this order. As $v(G) > l$ and G is connected, we may assume that the vertex $l+1 \in V(G)$ is connected to l. Obviously, we can add the edge $\{2, l+1\}$ which creates the l-cycle through $2, 3, \ldots, l, l+1$. Next, we can add the edge $\{1, 4\}$ which creates the l-cycle through $4, 5, \ldots, l, l+1, 2, 1$. Now the set $[l] \subset V(G)$ spans an l-cycle plus the edge $\{1, 4\}$ creating an odd $(l-2)$-cycle—the situation in which we can apply Theorem 49 to add all edges within $[l]$.

But it is trivial to show that a connected graph with an l-clique is weakly C_l-saturated, which implies $G \in \text{w-SAT}(n, C_l)$.

Now it is easy to deduce that w-SAT(n, C_l) consists exactly of all connected graphs containing a path of length $l-1$ as a subgraph and the desired characterization of the minimum ones follows.

Let us show that C_l admit a g-proof for $n > l$. Indeed, consider $\mathcal{G} = \mathcal{G}_{S_n^2}$. Any edge of C_l is $\mathcal{G}$-dependent on the remaining ones because the path with l edges is weakly K_3^2-saturated and K_3^2 is a $\mathcal{G}$-circuit. Clearly, $R_{\mathcal{G}}([n]^{(2)}) = n-1$. (In fact, if restricted to $[n]^{(2)}$, $\mathcal{G}$ is the usual *cycle matroid.*) The claim clearly follows. ∎

Theorem 51 *Let $l \geq 4$ be even and let $n \geq l$. Then* w-sat$(n, C_l) = n$*, all extremal graphs are trees of order n and diameter at least $l-1$ plus an extra edge creating an odd cycle, and C_l admits an r-proof for n.*

Proof. Similarly to the proof of Theorem 50, to show that any indicated graph G is weakly C_l-saturated, we first argue that adding a few edges we can obtain an l-cycle containing a 3-chord. Unfortunately, this configuration is not weakly C_l-saturated but, like in Theorem 49, we can add all 3-chords, 5-chords, and so on to obtain the complete bipartite graph $K_{l/2,l/2}$.

Observe that having an edge $\{x, y\}$ with y belonging be $K_{s,t}$-subgraph with $s, t \geq l/2$, we can connect x to any vertex lying in the same part as y. Hence, we can add edges so that G contains a $K_{s,n-s}$-subgraph with $s, n-s \geq l/2$; moreover, as we have an odd cycle present in the original G, one part spans an edge and $G \in$ w-SAT(n, C_l).

The required characterization of extremal graphs easily follows.

Finally, let $\mathcal{M}$ be Doob's [Doo73] *even-cycle matroid* on $[n]^{(2)}$ which can be represented by $f : [n]^{(2)} \to V$ which maps $\{i, j\}$ to $e_i + e_j$ for some basis $\{e_1, \ldots, e_n\}$ of a real vector space V. The cycle C_l is an $\mathcal{M}$-chain: if C_l goes via the vertices $1, 2, \ldots, l, 1$, then we have the linear relation

$$(e_1 + e_2) - (e_2 + e_3) + \ldots - (e_l + e_1) = 0$$

with all coefficients non-zero.

For $n \geq 3$, the rank of $\mathcal{M}$ is n as any basis vector e_i admits a representation $e_i = \frac{1}{2}((e_i + e_j) + (e_i + e_k) - (e_j + e_k))$, which implies our claim. ∎

Remark. Probably, even cycles do not admit a g-proof. But if we consider $(1,1)$-layered (i.e. bipartite) graphs, then C_4 as the complete $(1,1)$-graph admits a G-proof by the results of Section 12. A little more work shows that any even cycle admits a g-proof in the class of bipartite graphs, because any connected $(1,1)$-graph is weakly C_4-saturated.

Disjoint Edges

Suppose that we forbid lK_r^r, $l > 1$, that is, l disjoint r-edges.

Theorem 52 *Let $F = lK_r^r$, let $n > lr$, and let G consist of $l-1$ disjoint r-edges plus $n - r(l-1)$ isolated vertices. Then* $\text{w-sat}(n,F) = l-1$*, G is the only extremal graph, and the pair (F,G) admits a G-proof.*

Proof. Let us show that G is weakly F-saturated. As $v(G) > kl$, we can add an edge disjoint from the edges of G which creates a copy of F and leaves at least one vertex of G isolated.

Fix any $D \in [n]^{(r)}$. We have to show that $D \in \mathrm{Cl}_F(G)$. We prove that the existence of $E \in \mathrm{Cl}_F(G)$ with $|E \cap D| = k < r$ implies that there is $E' \in \mathrm{Cl}_F(G)$ with $|E' \cap D| = k+1$. Given E, there are $E_2, \ldots, E_l \in \mathrm{Cl}_F(G)$ which together with E form an F-subgraph. If there is $x \in D \setminus V$, $V = E \cup E_{[2,l]}$, then we can take $E' = E + x - y \in \mathrm{Cl}_F(G)$, for some $y \in E \setminus D$. Otherwise take any $x \in D \setminus E$, say $x \in E_2$, replace E_2 by $E_2' = E_2 - x + y \in \mathrm{Cl}_F(G)$, where $y \notin V$, and consider $E' = E - z + x$, $z \in E \setminus D$ which (together with $E_2', E_3, \ldots, E_l$) creates a forbidden subgraph. The required E' is found. Hence, $\text{w-sat}(n,F) \le l-1$.

Any weakly F-saturated graph contains $l-1$ disjoint edges; hence G is the only extremal graph.

The pair (F,G) admits a G-proof by Lemma 35. ∎

However, if the forbidden graph is a perfect matching, then the exact answer is known generally for $r = 2$ only.

Theorem 53 *For $n = 2l \ge 4$,* $\text{w-sat}(n, lK_2^2) = n-1$ *and all extremal graphs can be obtained in the following way: complete lK_2^2 to a tree T, add an edge E creating an odd cycle and remove any edge E' contained in some perfect matching of $T + E$.*

Proof. Let us show that any above constructed graph G is weakly F-saturated, $F = lK_2^2$. First we add the edge E'. Let C be the odd cycle (at this stage it is unique) of the obtained graph $T' = G + E' = T + E$ and let M be some matching of T'. Let $\{x_1, y_1\}, \dots, \{x_{2k+1}, y_{2k+1}\}$ be all edges of M with $x_i \in C$ and $y_i \notin C$.

Claim 1 A disjoint union of an odd cycle C_{2p-1} an a even path P_{2q} is weakly $(p+q)K_2^2$-saturated, $p \geq 2$, $q \geq 0$.

We prove the claim by induction on q. If $q = 0$ then we can first connect the isolated vertex to any other vertex of the cycle to obtain a wheel and then we can add the remaining edges in any order. If $q > 0$ then we can connect the endpoints of the path to all vertices on a cycle and the obtained graph is easily seen to contain $C_{2p+1} \sqcup P_{2q-2}$ and the claim follows.

A moment's thought reveals that, by Claim 1, $T' \in$ w-SAT(n, F) if $k = 0$. So, to prove that $G \in$ w-SAT(n, F), we show that, for $k > 0$, we can F-properly add some extra edges to T' and find other, strictly larger, odd cycle C.

Assume that $x_1, \dots, x_{2k+1}$ lie on the cycle C in this order clockwise. Note that we can add to T' all edges of the form $\{y_i, y_{i+1}\}$, $i \in [2k+1]$, which creates the matching $M' = M \triangle C_i$, where C_i the cycle via $y_{i+1}, y_i, x_i C x_{i+1}, y_{i+1}$ created just now. (By aCb we denote the part of the cycle C going clockwise from a to b inclusive.) If there are no vertices (along C) between some x_i and x_{i+1} then we have a strictly longer cycle $C \triangle C_i$ as desired. Otherwise, we may assume that a part of the cycle C looks like $x_1, \dots, a, x_2, b, \dots, c, x_3, \dots$. It is routine to check that the addition of the edge $\{a, c\}$ creates a matching which uses edges $\{x_2, b\}$, $\{y_1, y_2\}$, $\{x_i, y_i\}$, $i \in [3, 2k+1]$, etc. But then we can find a strictly longer odd cycle: $x_3 C a, c C^{-1} x_2, y_2, y_3, x_3$, which proves that $G \in$ w-SAT(n, F) as claimed.

On the other hand, consider any weakly F-saturated graph G and let $G' = G + E$ be a graph with a perfect matching. If G' is not connected, then all its components have even order, but then the first F-properly added edge not lying within a component cannot create a matching (by the parity argument), which is a contradiction. If G' is a bipartite graph, then its parts must be of the same size, but then the first F-properly added edge lying within a part creates no perfect matching, which a contradiction. Hence, G' is a tree plus an edge creating an odd cycle and all claims of the theorem easily follow. ∎

Dumb-Bells

Recall that the 2-graph B_{kk}, called a *dumb-bell*, consists of two disjoint copies of K_k^2 plus one edge connecting them, $k \geq 3$.

Theorem 54 *Let $k \geq 3$, $n = lk + q$, $q \in [0, k-1]$, $l \geq 2$; let $\varepsilon_q = 1$ except $\varepsilon_0 = 0$. Then* w-sat$(n, B_{kk}) = (l+1)\binom{k}{2} - \binom{k-q}{2} - \varepsilon_q$.

Proof. To prove the upper bound consider the 2-graph G on $[n]$ defined (for any q) by

$$E(G) = \left(\cup_{i\in[l]} A_i^{(2)}\right) \bigcup \left([n-k+1, n]^{(2)} \setminus \{n-1, n\}\right),$$

where $A_i = [ki - k + 1, ki]$, $i \in [l]$. As $G[kl] \cong lK_k^2$, we can add all missing edges within $[kl]$ each connecting some two of the A's. If $q = 0$, then we are done; otherwise we add the edge $\{n, n-1\}$ making $G[n-k+1, n]$ complete and then add the remaining edges in any order. Hence, $G \in$ w-SAT(n, B_{kk}) and the upper bound follows.

On the other hand, let $G \in$ w-SAT(n, B_{kk}) be arbitrary. Similarly to Lemma 43, we take a B_{kk}-proper ordering $\overline{G} = \{E_1, \dots, E_e\}$; assume that E_i creates a B_{kk}-subgraph F_i on a $2k$-set $M_i \subset [n]$. Define the *surplus* $s(X) = e(G[X]) - \frac{k-1}{2}|X|$, $X \subset [n]$, and $s_i = s(M_{[i]})$.

Let $q_i \in [0, k-1]$ be equal to $|M_i \setminus M_{[i-1]}| \pmod{k}$. Given q_i, it is routine to see that if $q_i = 0$ then $s_i \geq s_{i-1}$ and if $q_i > 0$ then

$$s_i - s_{i-1} \geq f(q_i) = \binom{k}{2} - \binom{k-q_i}{2} - \frac{k-1}{2}q_i - 1 \geq 0.$$

Furthermore, for $p, q \geq 1$, $f(p+q) \leq f(p) + f(q)$. Hence, $s(M_{[e]}) \geq f(q)$ for $q > 0$ and $s(M_{[e]}) \geq 0$ for $q = 0$. Now, the identity $e(G) = \frac{k-1}{2}n + s(V(G))$ implies the required lower bound. ∎

Remark. In fact, we c-prove that w-sat$(n, B_{kk}) \geq \frac{k-1}{2}n$ for odd k, which is sharp for $n = kl$. (For even k, the function $L = \frac{k-1}{2}p_1$ is not integral.)

Forests

Let us consider 2-graphs. Let T be a forest of order m. Clearly, K_{m-1}^2 plus $n - m + 1$ isolated vertices is weakly T-saturated, so w-sat$(n, T) \leq \binom{m-1}{2}$. This is sharp for $T = S^2_m$ by Corollary 38. The opposite extreme inequality is

$$\text{w-sat}(n, T) \geq e(T) - 1, \quad n \geq m. \tag{75}$$

By Lemma 35, if we have equality in (75), then T admits a g-proof for n. In fact, we can show that we have a G-proof.

Lemma 55 *Let F and H be any forests with $e(F) \leq e(H)$. Then F independent in $\mathcal{G}_H$.*

Proof. We use induction on $l = e(H)$. It is enough to prove the claim for $e(F) = e(H)$. Assume that 1 is an endvertex incident to the edge $E = \{1, 2\}$ in both F and H. Clearly,

$$\det(M(H, F)) = \pm \alpha_{1,1} \alpha_{2,2} \det(M(H - E, F - E)) + (\alpha_{1,1}\text{-free polynomial}).$$

By induction we conclude that $\det(M(H, F)) \neq 0$, which proves the lemma. ∎

Corollary 56 *If $G \in \text{w-SAT}(n, T)$, for some forest T, and $e(G) = e(T) - 1$ then the pair (T, G) admits a G-proof.*

Proof. Indeed, G is a forest. Also, T is dependent in $\mathcal{G}_G$ but, by Lemma 55, any proper subgraph of T is not; hence T is a $\mathcal{G}_G$-circuit. ∎

If T contains, for example, vertices a, b, c of degrees $1, 1, 2$ respectively such that $\{a, c\}, \{c, d\}, \{b, d\} \in E(T)$, for some vertex d, then adding the edges $\{d, x\}$ and $\{x, y\}$ to T, any $x, y \notin V(G)$, we create each time a new graph isomorphic to T; this implies equality in (75) with possible exceptions for some $n \leq 2m$. Generating a random tree by, for example, taking all m^{m-2} vertex-labelled trees equiprobable, one can show that almost every tree contains the above '*abc*-configuration' and therefore admits a G-proof.

The above results can be extended to *hypertrees*, for the definitions see Part IV, but we do not want to clutter the text with details.

11 Cones

In this section we prove that cones 'preserve' $G/g/\mathrm{g}'$-proofs. These results appear in [Pik99a].

To define the *cone* $\mathrm{cn}(G)$ of an r-graph G, add to G a new vertex and all edges containing this vertex. In other words, pick $v \notin V(G)$ and define $V(\mathrm{cn}(G)) = V(G) \cup \{v\}$ and

$$E(\mathrm{cn}(G)) = E(G) \cup \left\{ \{v\} \cup E : E \in V(G)^{(r-1)} \right\}.$$

For a family $\mathcal{F}$ of r-graphs, define $\mathrm{cn}(\mathcal{F}) = \{\mathrm{cn}(F) : F \in \mathcal{F}\}$.

For 2-graphs, $\mathrm{cn}^l(F) = K_l + F$; so, for example, the cones of empty graphs, cycles, complete graphs are stars, wheels and complete graphs respectively.

Lemma 57 *Suppose that every $r-1$ vertices of an r-graph F are covered by at least one edge. If F is a $\mathcal{G}_G$-chain, for some r-graph G, then $\mathrm{cn}(F)$ is a $\mathcal{G}_{\mathrm{cn}(G)}$-chain.*

Proof. Suppose first that $v(G) \geq v(F)$. Let $G' = \mathrm{cn}(G)$, $V(G) = [n-1]$ and $V(G') = [n]$. Identify the vertices of G' with the basis $\{e_1, \ldots, e_n\}$ of a vector space V'. Let Z' be the subspace of $\bigwedge^r V'$ and let $\mathcal{G}_{G'}$ be the gross matroid on $[n]^{(r)}$ corresponding to G'.

We may assume that $F' = \mathrm{cn}(F)$ is embedded into $[n]$ so that $V(F') \backslash V(F) = \{n\}$. We have to show that $E(F')$ is a chain in $\mathcal{G}_{G'}$, that is, we have to find $h' \in Z'$ such that $\mathrm{supp}(h') = E(F')$. Define $g_n^* = f_n^*$ and

$$g_i^* = f_i^* - \frac{\alpha_{in}}{\alpha_{nn}} f_n^*, \qquad i = 1, \ldots, n-1. \tag{76}$$

Recall that $\mathbf{f}^*$ is a generic basis of $(V')^*$ and $\alpha_{ij} = f_i^*(e_j)$, so $g_i^*(e_n) = 0$, $i \in [n-1]$, and this is the main point of our definition.

The matrix $N = (g_i^*(e_j))_{i,j \in [n-1]}$ is a generic matrix for a generic choice of the α's. Indeed, if its entries, $\gamma_{ij} = \alpha_{ij} - \alpha_{in}\alpha_{nj}/\alpha_{nn}$, $i,j \in [n-1]$, are algebraically dependent, then clearly the α's are.

As F is a $\mathcal{G}_G$-chain, the system of linear equations

$$g_D^* \llcorner \left(\sum_{E \in E(F)} c_E e_E \right) = 0, \qquad D \in E(G), \tag{77}$$

with respect to the undeterminants $(c_E)_{E \in E(F)}$, has a solution with all c's being non-zero for generic $\mathbf{g}$ (which is the case for generic $\mathbf{f}$). Apply elementary matrix transforms to write the system (77) in a diagonal form. For the free variables choose $\beta_1, \ldots, \beta_k$ which (together with the α's) are algebraically independent over the rationals and compute the other variables each being a rational function of the α's and β's.

Let $h = \sum_{E \in E(F)} c_E e_E$ and $h' = f_n^* \llcorner (h \wedge e_n)$. To complete the theorem it is enough to show that $h' \in Z'$ and $\mathrm{supp}(h') = E(F')$.

Let $D \in E(G')$. We want to show that $f_D^* \llcorner h' = 0$. If $D \ni n$, then

$$\langle f_D^*, h' \rangle = \langle f_{D \setminus \{n\}}^* \wedge f_n^*, f_n^* \llcorner (h \wedge e_n) \rangle = 0.$$

If $n \notin D$, that is, $D = \{d_1, \ldots, d_r\} \in E(G)$, then, by (76),

$$f_D^* = \bigwedge_{i=1}^r f_{d_i}^* = \bigwedge_{i=1}^r \left(g_{d_i}^* + \frac{\alpha_{d_i n}}{\alpha_{nn}} f_n^* \right) = g_D^* + f_n^* \wedge x^*,$$

some $x^* \in \bigwedge^{r-1} V^*$. Now,

$$\langle f_D^*, h' \rangle = \langle g_D^* + f_n^* \wedge x^*, f_n^* \llcorner (h \wedge e_n) \rangle = \langle g_D^* \wedge f_n^*, h \wedge e_n \rangle.$$

But for every $i \in [n-1]$ we have $g_i^*(e_n) = 0$, so the above expression is equal to $f_n^*(e_n) \langle g_D^*, h \rangle$ which is zero by the definition of h. Therefore $h' \in Z'$.

Let us show that $\mathrm{supp}(h') = E(F')$. Clearly, every $E \in \mathrm{supp}(h')$ either contains n or belongs to $E(F)$ which shows that $\mathrm{supp}(h') \subset E(F')$. On the other hand, take any $E \in E(F')$. If $E \in E(F)$, then

$$\langle e_E^*, h' \rangle = \langle e_E^* \wedge f_n^*, h \wedge e_n \rangle = \langle e_E^*, h \rangle \cdot \langle f_n^*, e_n \rangle = c_E f_n^*(e_n) \neq 0,$$

because $n \notin E$. If $E \ni n$, then let $D_1, \ldots, D_l$ be the edges of F containing $E' = E \setminus \{n\}$. By our assumption, $l > 0$. Let $D_i \setminus E = \{d_i\}$. Then

$$\begin{aligned} P_E &= \langle e_E^*, h' \rangle = \langle e_{E'}^* \wedge e_n^* \wedge f_n^*, h \wedge e_n \rangle = -\langle e_{E'}^* \wedge f_n^*, h \rangle \\ &= -\left\langle e_{E'}^* \wedge f_n^*, \sum_{E \in E(F)} c_E e_E \right\rangle = \sum_{i=1}^{l} \pm c_{D_i} \langle f_n^*, e_{d_i} \rangle = \sum_{i=1}^{l} \pm c_{D_i} \alpha_{n,d_i}. \end{aligned}$$

(The third equality is true as $\mathrm{supp}(h) = E(F) \subset [n-1]^{(r)}$.)

As every c_{D_i} is a rational function in the α's and β's, so is P_E. To show that $P_E \neq 0$ for a generic $\mathbf{f}$, it is enough to demonstrate an example of $\mathbf{f}$ when $P_E \neq 0$. Let $\alpha_{in} = 0$, $i \in [n-1]$. Then system (77) reduces to

$$f_D^* \llcorner \left(\sum_{E \in E(F)} c_E e_E \right) = 0, \qquad D \in E(G). \tag{78}$$

By the algebraic independence of $(f_i^*(e_j))_{i,j\in[n-1]}$, if we perform the diagonalisation for (78) in the same order as for (77), we will obtain the same set of free variables. Therefore, $(c_E)_{E\in E(F)}$ provides every solution for (78) when the β's range over the reals. Thus each c_E is non-zero (as F is a $\mathcal{G}_G$-chain) and it can depend only on $f_i^*(e_j)=\alpha_{ij}$, $i,j\in[n-1]$, and the β's. Now it is obvious that $P_E=\sum_{i=1}^{l} c_{D_i}\alpha_{n,d_i}$ cannot be identically zero. This proves the lemma if $v(G)\ge v(F)$.

Otherwise, we can add $v(F)-v(G)$ isolated vertices to G to obtain H. By above, $\mathrm{cn}(F)$ is a chain in $\mathcal{G}_{\mathrm{cn}(H)}$, that is, each edge of $\mathrm{cn}(F)$ is dependent on the other edges. The latter claim is certainly true in $\mathcal{G}_{\mathrm{cn}(G)}$ which has more dependences than $\mathcal{G}_{\mathrm{cn}(H)}$ as $\mathrm{cn}(G)\subset\mathrm{cn}(H)$. ∎

Lemma 58 *If an r-graph F is independent in $\mathcal{G}_G$ and $v(F)\le v(G)$, then* $\mathrm{cn}(F)$ *is independent in $\mathcal{G}_{\mathrm{cn}(G)}$.*

Proof. We assume the same conventions as those appearing in the proof of Lemma 57 before (77).

It is enough to prove our claim in the case $e(G)=e(F)$: if $e(G)>e(F)$ we can remove a G-edge with F being still $\mathcal{G}_G$-independent.

Let us show that the rank of $M'(G',F')$ is $e(F')$, where $M'(D,E)=\langle g_D^*,e_E\rangle$, $D\in E(G')$, $E\in E(F')$, which would imply the lemma.

By our assumption, the square submatrix $M'(G,F)\subset M'(G',F')$ is non-singular because the matrix N is generic. As $g_i^*(e_n)=0$ for $i\in[n-1]$, we conclude that all entries of the submatrix $M'(G,F'')$ are zeros, where $E(F'')=E(F')\setminus E(F)$. Therefore, to prove the claim we have to show that the submatrix $M'(G'',F'')$ has the maximal possible rank $\binom{v(F)}{r-1}$, where $E(G'')=E(G')\setminus E(G)$.

For any $D'=D\cup\{n\}\in E(G'')$, $E'=E\cup\{n\}\in E(F'')$, we have

$$\langle g_{D'}^*,e_{E'}\rangle=g_n^*(e_n)\cdot\langle g_D^*,e_E\rangle,$$

because $g_i^*(e_n)=0$, $i\in[n-1]$. (As n is the last element in D' and E', we do not have ±1 in the formula.) Now,

$$M'(G'',F'')=g_n^*(e_n)\cdot M'\big(K^{r-1}([n-1]),K^{r-1}(V(F))\big)$$

has rank $\binom{v(F)}{r-1}$ because N is generic. ∎

Remark. Is is not hard to show that if F is not independent in $\mathcal{G}_G$, then $\mathrm{cn}(F)$ is not independent in $\mathcal{G}_{\mathrm{cn}(G)}$, for any r-graphs F and G. But we do not need this result.

Lemma 59 *If $G \in$ w-SAT$(n-1, \mathcal{F})$, then $\mathrm{cn}(G) \in$ w-SAT$(n, \mathrm{cn}(\mathcal{F}))$. In particular,*

$$\text{w-sat}(n, \mathrm{cn}(\mathcal{F})) \le \text{w-sat}(n-1, \mathcal{F}) + \binom{n-1}{r-1}.$$

Proof. Let $E_1, \ldots, E_m$ be an $\mathcal{F}$-proper ordering of $E(\overline{G})$. To show that $G' = \mathrm{cn}(G)$ is weakly $\mathrm{cn}(\mathcal{F})$-saturated, add these edges in the same order to G'. (Note that $E(\overline{G'}) = E(\overline{G})$.) Every E_i creates an F-subgraph in G, $F \in \mathcal{F}$, which, together with the extra vertex, creates a copy of $\mathrm{cn}(F)$ in G', so $G' \in$ w-SAT$(n, \mathrm{cn}(\mathcal{F}))$. ∎

Theorem 60 *Let $\mathcal{F}$ be a family of r-graphs such that in each $F \in \mathcal{F}$ every $r-1$ vertices are covered by at least one edge.*

If a pair $(\mathcal{F}, G)$ admits a G-proof, then the pair $(\mathrm{cn}(\mathcal{F}), \mathrm{cn}(G))$ admits a G-proof.

If we can g-prove w-sat$(n-1, \mathcal{F}) \ge l$, then we can g-prove

$$\text{w-sat}(n, \mathrm{cn}(\mathcal{F})) \ge l + \binom{n-1}{r-1}. \tag{79}$$

In particular, if $\mathcal{F}$ admits a g-proof for $n-1$, then $\mathrm{cn}(\mathcal{F})$ admits a g-proof for n. The analogous claim is true for the g'-technique.

Proof. Let us consider G-proofs first. By Lemma 59, $\mathrm{cn}(G)$ is weakly $\mathrm{cn}(\mathcal{F})$-saturated. By Lemma 57, $\mathrm{cn}(F)$ is a $\mathcal{G}_{\mathrm{cn}(G)}$-chain for every $F \in \mathcal{F}$. Hence, the pair $(\mathrm{cn}(\mathcal{F}), \mathrm{cn}(G))$ admits a G-proof.

Next, consider the g-technique. Take any G such that each $F \in \mathcal{F}$ is a $\mathcal{G}_G$-chain and $R_{\mathcal{G}_G}(K_{n-1}^r) \ge l$. Adding extra vertices to G, we may assume $v(G) \ge n-1$. By Lemma 57, each graph in $\mathrm{cn}(\mathcal{F})$ is a chain in $\mathcal{G}_{\mathrm{cn}(G)}$.

By definition, $R_{\mathcal{G}_G}(K_{n-1}^r) \ge l$, so choose a $\mathcal{G}_G$-independent subgraph $H \subset K_{n-1}^r$ of rank l. Assume $v(H) = n-1$. By Lemma 58, $\mathrm{cn}(H)$ is independent in $\mathcal{G}_{\mathrm{cn}(G)}$. Hence, the rank of K_n^r in $\mathcal{G}_{\mathrm{cn}(G)}$ is at least $e(\mathrm{cn}(H)) = l + \binom{n-1}{r-1}$, that is, we can g-prove (79), as required.

In the g'-case, choose G such that each $F \in \mathcal{F}$ is a $\mathcal{G}_G$-chain and

$$R_{\mathcal{G}_G}(K_{n-1}^r) + D_{\mathcal{G}_G}(\mathcal{F}) - 1 \geq l.$$

Now we proceed in the same way as in the g-case, except we have to show additionally that, for any $F \in \mathcal{F}$, we have $D_{\mathcal{G}_G}(F) \leq D_{\mathcal{G}_{\mathrm{cn}(G)}}(\mathrm{cn}(F))$.

Note that if we have F-edges $E_1, \ldots, E_d$ whose removal does not decrease the $\mathcal{G}_G$-rank of $E(F)$, then the system of equations (77) has a solution in which $c_{E_1}, \ldots, c_{E_d}$ can be chosen to be the free variables $\beta_1, \ldots, \beta_d$. Following the proof of Lemma 57 (note that F is a $\mathcal{G}_G$-chain), one can let (c_E) be such a solution of (77) and observe that

$$\langle e_{E_i}^*, h' \rangle = \langle e_{E_i}^* \wedge f_n^*, h \wedge e_n \rangle = \langle e_{E_i}^*, h \rangle \cdot \langle f_n^*, e_n \rangle = \beta_i \alpha_{nn}, \quad i \in [d],$$

since $E_i \subset [n-1]$. This means that, choosing generic β's, we can obtain $h' \in Z'$ whose support is $E(\mathrm{cn}(F))$ with $e_{E_i}^*(h')$ being generic, which is precisely to say that $E_1, \ldots, E_d$ are $\mathcal{G}_{\mathrm{cn}(G)}$-dependent on the other edges of $\mathrm{cn}(F)$. Hence, $D_{\mathcal{G}_{\mathrm{cn}(G)}}(\mathrm{cn}(F)) \geq d$ and the claim follows. ∎

Remark. We cannot generally discard the covering condition in Lemma 57 or Theorem 60. (But note that we do not have any covering condition on G.) Consider, for example, $r = 2$ when the condition rules out isolated vertices. Let F be a triangle plus an isolated vertex and let G be a star $K_{1,n-2}$, $n \geq 5$. Then (F, G) admits a G-proof (see Subsection 10.1). But it is easy to see that $\text{w-sat}(n, \mathrm{cn}(F)) = 6 < e(\mathrm{cn}(G)) = 2n - 3$, and so $\mathrm{cn}(F)$ cannot be a $\mathcal{G}_{\mathrm{cn}(G)}$-chain.

We noted already in Section 7 that many new results can be proved by applying Theorem 60, so we do not repeat these examples here.

12 Joins

Here we indicate how to extend the idea of G/g/etc.-proof to layered graphs (which were defined in Subsection 4.2) and prove that *joins* 'preserve' G/g/r-proofs. These results appear in [Pik99a].

The notion of weak saturation extends to layered graphs in the obvious way. For example, given an $\mathbf{r}$-graph $\mathbf{F}$, $\text{w-SAT}(\mathbf{n}, \mathbf{F})$ consists of all $\mathbf{r}$-graphs $\mathbf{G}$ on an

$\mathbf{n}$-set such that we can consecutively add all missing $\mathbf{r}$-edges to $\mathbf{G}$ creating every time an $\mathbf{F}$-subgraph.

It is clear how to extend the notion of an m/r-proof to layered graphs. It is possible also to introduce the gross matroid of an $\mathbf{r}$-graph $\mathbf{G}$ defined on an $\mathbf{n}$-set $\mathbf{X}$. Indeed, identify each X_i with a basis $\mathbf{e}_i = (e_{i,j})_{j\in[n_i]}$ of an n_i-dimensional vector space V_i and consider $\bigwedge \mathbf{V}$ which, by the definition, is the tensor product of the exterior algebras over V_i, $i \in [t]$:

$$\bigwedge \mathbf{V} = \bigotimes_{i\in[t]} \bigwedge V_i.$$

Let $\bigwedge^{\mathbf{r}} \mathbf{V}$ be the linear subspace of $\bigwedge \mathbf{V}$ spanned by the elements

$$h = h_1 \otimes \ldots \otimes h_t, \qquad h_i \in \bigwedge^{r_i} V_i, \quad i \in [t].$$

Let $\mathbf{f}_i = (f_{i,j})_{j\in[n_i]}$ be another basis of V_i lying in generic position with respect to $\mathbf{e}_i$, $i \in [t]$.

In the obvious way we define supports, etc. For any $\mathbf{r}$-subset $\mathbf{E} \subset \mathbf{X}$, let

$$\mathbf{f}_{\mathbf{E}} = \bigotimes_{i\in[t]} f_{i,E_i} \quad \text{and} \quad \mathbf{e}_{\mathbf{E}} = \bigotimes_{i\in[t]} e_{i,E_i},$$

or, in other words, in every $\bigwedge V_i$, we take the element corresponding to E_i in the basis $\mathbf{f}_i$ or $\mathbf{e}_i$ and then compute the tensor product. Let the linear subspace $\mathbf{Z} \subset \bigwedge^{\mathbf{r}} \mathbf{V}$ corresponding to $\mathbf{G}$ be spanned by the elements $\{\mathbf{f}_{\mathbf{E}} : \mathbf{E} \in E(\overline{\mathbf{G}})\}$ and let $\mathbf{r}$-sets $\mathbf{E}_1, \ldots, \mathbf{E}_k$ be independent if no linear combination of $\mathbf{e}_{\mathbf{E}_1}, \ldots, \mathbf{e}_{\mathbf{E}_k}$ (except 0) belongs to $\mathbf{Z}$. The required matroid $\mathcal{G}_{\mathbf{G}}$ of rank $\operatorname{codim}(\mathbf{Z}) = e(\mathbf{G})$ is built. Clearly, it is *symmetric*, that is, invariant under layer-preserving permutations.

Given t (usual) r_i-graphs F_i, $i \in [t]$, with disjoint vertex sets, their *join* (or *tensor product*) $\mathbf{F} = F_1 \otimes \ldots \otimes F_t$ is the layered $\mathbf{r}$-graph on the layered set $V(\mathbf{F}) = (V(F_1), \ldots, V(F_t))$ such that an $\mathbf{r}$-subset $\mathbf{E} = (E_1, \ldots, E_t)$ is an edge of $\mathbf{F}$ if and only if $E_i \in E(F_i)$ for every $i \in [t]$. Thus $e(\mathbf{F}) = \prod_{i\in[t]} e(F_i)$. For example, the join of two 1-graphs is a complete bipartite graph (possibly plus isolated vertices).

Suppose that we are given t families $\mathcal{F}_i$ of r_i-graphs, $i \in [t]$. We define their *join* by

$$\mathcal{F} = \mathcal{F}_1 \otimes \ldots \otimes \mathcal{F}_t = \{F_1 \otimes \ldots \otimes F_t : F_i \in \mathcal{F}_i,\ i \in [t]\}.$$

Let these conventions apply to the following results.

Lemma 61 *If $G_i \in$ w-SAT$(n_i, \mathcal{F}_i)$, $i \in [t]$, then $\mathbf{G} \in$ w-SAT$(\mathbf{n}, \mathcal{F})$, where $\overline{\mathbf{G}} = \overline{G_1} \otimes \ldots \otimes \overline{G_t}$. In particular,*

$$\text{w-sat}(\mathbf{n}, \mathcal{F}) \le \prod_{i\in[t]} \binom{n_i}{r_i} - \prod_{i\in[t]} \left(\binom{n_i}{r_i} - \text{w-sat}(n, \mathcal{F}_i) \right).$$

Proof. Denote $b_i = e(\overline{G_i})$. Let $E_{i,j} \in \overline{G_i}$, $j = 1, \ldots, b_i$, be an $\mathcal{F}_i$-proper ordering, $i \in [t]$. There is the obvious bijective correspondence between the elements in $B = [b_1] \times \ldots \times [b_t]$ and the edges of $\overline{\mathbf{G}}$ which maps $(j_1, \ldots, j_t)$ to $\cup_{i\in[t]} E_{i,j_i}$.

Now we add the missing edges to $\mathbf{G}$ so that the corresponding elements of B are taken in the lexicographic order. Consider any added edge $\mathbf{E}$. Let $H_i \subset X_i$ be an F_i-subgraph created by E_i. (Note that $E_i \not\in E(G_i)$ by the definition of $\mathbf{G}$.) We claim that $\mathbf{H} = H_1 \otimes \ldots \otimes H_t$ is a forbidden subgraph created by $\mathbf{E}$. Indeed, let $\mathbf{D} \neq \mathbf{E}$, be an edge of $\mathbf{H}$. Clearly, for each $i \in [t]$, the edge $D_i \in E(H_i)$ must be present in G_i or be added before E_i or equal to E_i. If $D_i \in E(G_i)$ for at least one index i then $\mathbf{D} \in E(\mathbf{G})$. If not, then clearly the edge $\mathbf{D}$ comes before $\mathbf{E}$, as required.

Finally, $e(\overline{\mathbf{G}}) = \prod_{i\in[t]} e(\overline{G_i})$, which completes the proof. ∎

Lemma 62 *If F_i is a chain in $\mathcal{G}_{G_i}$, $i \in [t]$, then $\mathbf{F} = F_1 \otimes \ldots \otimes F_t$ is a chain in $\mathcal{G}_{\mathbf{G}}$, where $\overline{\mathbf{G}} = \overline{G_1} \otimes \ldots \otimes \overline{G_t}$.*

Proof. By the assumption, there is $h_i \in Z_{G_i} \subset \bigwedge^{r_i} V_i$ such that $\text{supp}_{\mathbf{e}_i}(h_i) = E(F_i)$, $i \in [t]$. Consider

$$h = h_1 \otimes \ldots \otimes h_t \in \textstyle\bigwedge^{\mathbf{r}} \mathbf{V}.$$

Obviously, $\text{supp}_{\mathbf{e}}(h) = E(\mathbf{F})$ and $\text{supp}_{\mathbf{f}}(h) \subset E(\overline{\mathbf{G}})$. Therefore, $h \in Z_{\mathbf{G}}$ and every edge in $\mathbf{F}$ is dependent on the rest, as required. ∎

Theorem 63 *Suppose that, for every $i \in [t]$, the pair $(\mathcal{F}_i, G_i)$ admit a G-proof. Then so does the pair $(\mathcal{F}, \mathbf{G})$, where $\overline{\mathbf{G}} = \overline{G_1} \otimes \ldots \otimes \overline{G_t}$.*

Suppose that, for each $i \in [t]$, we can g-prove that w-sat$(n_i, \mathcal{F}_i) \ge l_i$. *Then we can g-prove that*

$$\text{w-sat}(\mathbf{n}, \mathcal{F}) \ge \prod_{i\in[t]} \binom{n_i}{r_i} - \prod_{i\in[t]} \left(\binom{n_i}{r_i} - l_i \right). \tag{80}$$

In particular, if each $\mathcal{F}_i$ admits a g-proof for n_i, then $\mathcal{F}$ admits a g-proof for $\mathbf{n}$*. The analogous statement is true for the r-technique.*

Proof. Let us consider G-proofs first. By Lemma 61, $\mathbf{G} \in$ w-SAT$(\mathbf{n}, \mathcal{F})$, and by Lemma 62, every $F_1 \otimes \ldots \otimes F_k \in \mathcal{F}$ is a $\mathcal{G}_{\mathbf{G}}$-chain, and the claim follows.

Now, consider the g-case. For $i \in [t]$, choose G_i such that each graph in $\mathcal{F}_i$ is a $\mathcal{G}_{G_i}$-chain and the $\mathcal{G}_{G_i}$-rank of $K_{n_i}^{r_i}$ is at least l_i; let $H_i \subset K_{n_i}^{r_i}$ be a $\mathcal{G}_{G_i}$-independent subgraph of size l_i and order n_i. Let

$$\begin{aligned} \overline{\mathbf{G}} &= \overline{G_1} \otimes \ldots \otimes \overline{G_t}, \\ \overline{\mathbf{H}} &= \overline{H_1} \otimes \ldots \otimes \overline{H_t}. \end{aligned}$$

By Lemma 62, each $F_1 \otimes \ldots \otimes F_k \in \mathcal{F}$ is a $\mathcal{G}_{\mathbf{G}}$-chain.

Let us show that $\mathbf{H}$ is independent in $\mathcal{G}_{\mathbf{G}}$. As each H_i is $\mathcal{G}_{G_i}$-independent, we can find a linear map $p_i : \bigwedge^{r_i} V_i \to Z_{G_i}$ which is the identity map on Z_{G_i} while $p_i(e_E) = 0$ if $E \in E(H_i)$, $i \in [t]$. Define

$$\mathbf{p} = p_1 \otimes \ldots \otimes p_t : \bigwedge^{\mathbf{r}} \mathbf{V} \to Z_{G_1} \otimes \ldots \otimes Z_{G_t},$$

that is $\mathbf{p}(u_1 \otimes \ldots \otimes u_t) = p_1(u_1) \otimes \ldots \otimes p_t(u_t)$. Now, $\mathbf{p}$ is the identity map on $Z_{G_1} \otimes \ldots \otimes Z_{G_t} = Z_{\mathbf{G}}$, while $\mathbf{p}$ is zero on $e_{\mathbf{E}}$ for each $\mathbf{E} = E_1 \cup \ldots \cup E_t \in E(\mathbf{H})$: $E_i \in E(H_i)$ for some $i \in [t]$ and then $p_i(e_{E_i}) = 0$. Hence, no non-zero linear combination of $e_{\mathbf{E}}$, $\mathbf{E} \in E(\mathbf{H})$ can lie in $Z_{\mathbf{G}}$, that is, $\mathbf{H}$ is independent in $\mathcal{G}_{\mathbf{G}}$. The size of $\mathbf{H}$ equals the right-hand side of (80), as required.

The claim about g-proofs follows from Lemma 61.

In the r-case, our task is to construct a matroid $\mathcal{M}$ on the set of $\mathbf{r}$-subsets of $\mathbf{X}$ such that every graph in $\mathcal{F}$ is an $\mathcal{M}$-chain, should we be given appropriate matroids $\mathcal{M}_i$ on $Y_i = X_i^{(r_i)}$, $i \in [t]$.

Let $k_i : Y_i \to V_i$, for some vector space V_i, be a representation of the matroid $\mathcal{M}_i$, $i \in [t]$. Identify Y_i with a basis of some vector space W_i via $g_i : Y_i \hookrightarrow W_i$. Let $h_i : W_i \to V_i$ be the linear map extending k_i. Denote $Z_i = \ker(h_i) \subset W_i$. Clearly, $\operatorname{codim} Z_i = R_{\mathcal{M}_i}(Y_i) = e(G_i) \geq l_i$, where G_i is a base of $\mathcal{M}_i$.

Let $\overline{\mathbf{G}} = \overline{G_1} \otimes \ldots \otimes \overline{G_t}$. Identify the $\mathbf{r}$-subsets of $V(\mathbf{G})$ with a basis of $\mathbf{W} = \bigotimes_{i\in[t]} W_i$ by mapping $\mathbf{E} = (E_1, \ldots, E_t)$ into $g(\mathbf{E}) = \bigotimes_{i\in[t]} g_i(E_i)$. Let $\mathbf{Z} = \bigotimes_{i\in[t]} Z_i \subset \mathbf{W}$ and $p : \mathbf{W} \to \mathbf{W}/\mathbf{Z}$ be the projection.

Let $\mathcal{M}$ be the matroid represented by $p \circ g : V(\mathbf{G})^{(\mathbf{r})} \to \mathbf{W}/\mathbf{Z}$. Let us show that $\mathcal{M}$ r-proves (80).

As $g(V(\mathbf{G})^{(\mathbf{r})})$ is a basis for $\mathbf{W}$, we conclude that the rank of $\mathcal{M}$ is

$$\dim \mathbf{W} - \dim \mathbf{Z} = \prod_{i \in [t]} \binom{n_i}{r_i} - \prod_{i \in [t]} e(\overline{G_i}),$$

which is at least the right-hand side of (80).

Thus, all we have to do is to check that any $\mathbf{F} = F_1 \otimes \ldots \otimes F_t \in \mathcal{F}$ is an $\mathcal{M}$-chain. Fix an edge $\mathbf{E} = (E_1, \ldots, E_t) \in E(\mathbf{F})$. As F_i is an $\mathcal{M}_i$-chain, we conclude that there are $c_{i,E} \in \mathbb{R}$, $E \in E(F_i) \setminus \{E_i\}$, and $z_i \in Z_i$ such that

$$g_i(E_i) = z_i + \sum_{D \in E(F_i) \setminus \{E_i\}} c_{i,D}\, g_i(D), \qquad i \in [t]. \tag{81}$$

If we take the tensor product of (81) over $i \in [t]$, we obtain on the left-hand side the element $g(\mathbf{E})$ while on the right-hand side we will have $z_1 \otimes \ldots \otimes z_t \in \mathbf{Z}$ plus some other tensor products. Next, in the remaining tensor products replace each z_i by the linear combination of $(g_i(D))_{D \in E(F_i)}$ by (81). Each term then becomes $\otimes_{i \in [t]} g_i(D_i)$ for some $D_i \in E(F_i)$, i.e., it is of the form $g(\mathbf{D})$, $\mathbf{D} = (D_1, \ldots, D_t) \in E(\mathbf{F})$ and, moreover, we never have $\mathbf{D} = \mathbf{E}$. So we have a representation of $g(\mathbf{E})$ as a linear combination of an element of $\mathbf{Z}$ and of $g(\mathbf{D})$, $\mathbf{D} \in E(\mathbf{F}) \setminus \{\mathbf{E}\}$ which is precisely the required. The theorem is proved. ∎

Unfortunately, there does not seem to be a natural tensor product operation for matroids, cf. Lovász [Lov77], so we do not know if joins preserve m-proofs.

Alon [Alo85] (a different proof is presented by Yu [Yu93]) solved one extremal problem for set systems, which can be easily seen equivalent to computing the w-sat-function for joins of complete graphs. As complete graphs admit a G-proof (e.g. by Theorem 60), the result of Alon can be deduced as a special case of Theorem 63.

Part III
Chain Decompositions

13 Introduction

13.1 Discussion

There are many important results about *chain decompositions* of posets, that is, collections of chains such that every element in the poset belongs to exactly one chain. (We will also refer to these as *vertex decompositions.*) Typical questions are the following. What is the minimal number of chains of such a partition? Do there exist partitions with some extra properties (e.g. into symmetric chains)? Are there any applications of these decompositions?

In this part we investigate the notion of an *edge decomposition* which is a collection of chains such that every pair of adjacent elements (one covers the other) belongs to exactly one chain and we try to answer the above questions.

Such considerations may arise, for example, when in a computer programme we want to operate with posets, and so we wish to represent them efficiently in the memory. If keeping the relational binary $n \times n$-table is impossible or undesirable, we can try to maintain a list of chains completely determining the poset, and a natural question to ask is, for example, how small such a list can be. The related notion of *line poset* also arises naturally.

In Section 14 we compute the minimal size of a skipless chain decomposition of a poset in terms of other parameters, which can be viewed as an analogue of Dilworth's theorem [Dil50]. Surprisingly, this fundamental theorem is a new result. We prove it using the linear programming method of Dantzig and Hoffman [DH56]. Graham Brightwell simplified our proof by replacing the linear programming argument by an easy application of Hall's theorem. We present both these proofs.

The minimal size of an edge decomposition of $\mathcal{P}$ can be deduced as a corollary but we present a short and direct proof.

In Section 15 we provide an explicit edge decomposition of the lattice of subsets of a finite set into symmetric chains. Although the existence of such a partition can be deduced from the results of Anderson [And67] and Griggs [Gri77], a constructive proof seems to be unknown. The discovered partition has some extra properties and interesting applications. For the latter we refer the reader to Section 16.

In Section 17 we characterize line posets in terms of forbidden configurations and point out which information determines and can be reconstructed from its line poset.

13.2 Definitions

Let $\mathcal{P} = (X, >)$ be a poset (a partially ordered set). We say that y *covers* x (denoted by $y \gtrdot x$ or $x \lessdot y$) if $y > x$ and no $z \in X$ satisfies $x < z < y$ (such x, y will be also called *adjacent elements*). With every poset $\mathcal{P}$ we associate its *Hasse diagram* $D = D(\mathcal{P})$ which is the digraph with X as the vertex set and $(x, y) \in E(D)$ iff y covers x. Given a cycle-free digraph D, we can build a poset on the same vertex set by letting $x < y$ if there is a directed xy-path. Note that a cycle-free digraph D is the Hasse diagram of some $\mathcal{P}$ if and only if for every $(x, y) \in E(D)$ there is no directed xy-path of length greater than 1. The correspondence 'posets-digraphs' is very useful, so we often switch between the poset and digraph terminology without any warning.

A chain in $\mathcal{P}$ is called *skipless* if every element covers its predecessor; skipless chains correspond to *oriented paths* in the Hasse diagram. The *width* $w(\mathcal{P})$ is the maximal size of an antichain in $\mathcal{P}$.

The *line poset* $L(\mathcal{P})$ of a poset $\mathcal{P}$ has as the vertex set the pairs (x, y) of elements in $\mathcal{P}$ with y covering x and we agree that $(x \lessdot y)$ is less than $(x' \lessdot y')$ in $L(\mathcal{P})$ if and only if $y \le x'$. (This operation somewhat resembles taking the line graph, hence the name.)

Every skipless chain in $\mathcal{P}$ corresponds to a skipless chain in $L(\mathcal{P})$ of size smaller by 1. We usually identify these chains.

One can ask which important poset properties are preserved by the operator L. In fact, L preserves very few properties (e.g. self-duality, regularity). As in almost every case it is trivial to find a counterexample/proof, we do not dwell on this topic.

A *vertex partition* (*decomposition*) of $\mathcal{P}$ is a collection of chains such that every $x \in X$ belongs to exactly one chain. An *edge partition* (*decomposition*) is a family of skipless chains such that every pair $x, y \in X$ with x being covered by y belongs to exactly one chain. Note that the chains in an edge decomposition are required to be skipless. One can see that edge partitions of $\mathcal{P}$ correspond to vertex partitions of $L(\mathcal{P})$ into skipless chains.

The subsets of $[n]$ partially ordered via the inclusion relation, form the ranked poset $\mathcal{B}_n = (2^{[n]}, \subset)$. The corresponding Hasse diagram is the *oriented n-cube* Q_n. For $\mathcal{B}_n$, the relation 'B covers A' is denoted by $A \sqsubset B$.

We find it useful to identify $A \in \mathcal{B}_n$ with its *()-representation* which is the n-sequence of left and right parentheses corresponding to the elements of $\overline{A} = [n] \setminus A$ and A respectively. Likewise, the *$(*)$-representation* of an element $(A \sqsubset B) \in L(\mathcal{B}_n)$ contains '(' for the elements in $\overline{B}$, ')' for the elements in A and '$*$' for the element in $B \setminus A$.

Generally, let F be a sequence containing left and right parentheses. Consecutively and as long as possible remove matched pairs of adjacent brackets, ie. substrings '( )'. (Clearly, the order of operations does not matter.) The elements which would be removed by the above *matching* are called *fixed* or *paired* elements and the remaining ones are called *free*. In particular, the free parentheses always form the following (possibly empty) sequence:)) . . .)) ((. . . ((.

14 Skipless Chain Decompositions

Here we present a theorem computing the minimal number of skipless chains partitioning a given poset P. In fact, we prove a more general result about directed graphs.

Let D be any digraph. We may have loops and may have edges (i, j) and (j, i) simultaneously. Consider partitions of $V(D)$ into vertex-disjoint directed cycles and directed paths. (We consider any isolated vertex as a path of length zero; loops and pairs of opposite edges are considered as cycles.) Let $m(D)$ be the minimal number of directed paths in a such partition.

On the other hand, let $M(D)$ be the maximal value of $|A| - |B|$ taken over all pairs of disjoint sets $A, B \subset V(D)$ such that any directed path connecting two distinct vertices from A, contains a vertex of B and any cycle intersecting A intersects B. (In particular, if $(i, i) \in E(D)$ then $i \notin A$.) Clearly, for any such pair (A, B) we have $|P \cap A| \leq |P \cap B| + \varepsilon$, where $\varepsilon = 1$ if P is a directed path and $\varepsilon = 0$ if P is a directed cycle. This implies that $m(D) \geq M(D)$.

We will show that we have in fact equality for any D. Our proof is a modification of the proof by Dantzig and Hoffman [DH56] of Dilworth's theorem, which exploits methods of linear programming. (A simpler argument by Graham Brightwell is outlined after our proof.)

Theorem 64 *For any directed graph D we have $m(D) = M(D)$.*

Proof. As we have already observed $m(D) \geq M(D)$, so let us prove the converse inequality. Assume that $V(D) = [n]$. For $i, j \in [n]$ define $c_{00} = 1$, $c_{0j} = c_{i0} = 0$, and

$$c_{ij} = \begin{cases} 0, & \text{if } (i,j) \in E(D), \\ -\infty, & \text{if } (i,j) \notin E(D). \end{cases}$$

Consider the linear programming problem of finding k, where

$$k = \max \sum_{i,j \in [0,n]} c_{ij} x_{ij}, \tag{82}$$

given the following restrictions:

$$\sum_{j=0}^{n} x_{0j} = \sum_{i=0}^{n} x_{i0} = n, \tag{83}$$

$$\sum_{j=0}^{n} x_{ij} = \sum_{j=0}^{n} x_{ji} = 1, \qquad i \in [n], \tag{84}$$

$$x_{ij} \geq 0, \qquad i, j \in [0,n]. \tag{85}$$

Restrictions (83), (84) and (85) define a non-empty set; for example, we can satisfy them by letting x_{ij} be 0 for $i, j \in [n]$ and 1 otherwise, except $x_{00} = 0$. As for any feasible solution we have $x_{00} \leq n$ while the coefficients c_{ij} at other variables are non-positive, we conclude that the right-hand side of (82) is at most n and thus k is well-defined.

We claim that we can choose an integral solution to (82), that is, we can ensure that each x_{ij} is an integer. To do so, take a solution in which as many as possible variables are integers. Suppose there is $x_{i_1 i_2} \notin \mathbb{Z}$. By (83) or (84), the i_2th column contains another non-integer, $x_{i_3 i_2}$. Next, we consider the i_3th row and find $x_{i_3 i_4} \notin \mathbb{Z}$, and so on, until considering a current variable $x_{i_s i_t}$ we have a chance to select a previously chosen variable $x_{i_u i_v}$. In fact, we may have two choices at this step, but we will always be able to chose one with $s+v$ and $t+u$ being odd. Then the subsequence S of elements between $x_{i_s i_t}$ and $x_{i_u i_v}$ (inclusive) is of even length. If we add any ε to each $x_{i_k i_{k+1}} \in S$ and subtract ε from each $x_{i_{k+1} i_k} \in S$, then we do not affect (83) and (84). (Because each row or column contains either two variables, which receive different signs, or none.) The function $\sum c_{ij} x_{ij}$ is linear in ε, suppose it is non-decreasing. Let ε be the minimum of the fractional part of $x_{i_{k+1} i_k} \in S$; then our transformation makes at least one more variable integral, while (85) still holds. This contradiction proves the claim.

Any x_{ij}, except perhaps x_{00}, is either 0 or 1. A moment's thought reveals that by (84) the set $\{(i,j) \in [n]^2 : x_{ij} = 1\} \subset E(D)$ is a union of vertex-disjoint directed paths (this is to include isolated vertices) and cycles partitioning $V(D)$. The number of paths equals the number of occurrences of 1 among x_{0j}, $j \in [n]$, which by (83) is $n - x_{00} = n - k$. Hence,

$$m(D) \leq n - k. \tag{86}$$

Now, the Duality Theorem asserts that

$$k = \min \left(n(u_0 + v_0) + \sum_{i=1}^{n} u_i + \sum_{j=1}^{n} v_j \right), \tag{87}$$

given the following restrictions on variables u_i, v_i, $i \in [0,n]$:

$$u_0 + v_0 \geq 1, \tag{88}$$

$$u_i + v_0 \geq 0, \qquad i \in [n], \tag{89}$$

$$u_0 + v_j \geq 0, \qquad j \in [n], \tag{90}$$

$$u_i + v_j \geq 0, \qquad (i,j) \in E(D). \tag{91}$$

We claim that we can choose an integral solution to (87). To do so, take a solution with as many as possible variables among u_i, v_i, $i \in [0,n]$, being integers. Let $I = \{i \in [0,n] : u_i \notin \mathbb{Z}\}$ and $J = \{j \in [0,n] : v_j \notin \mathbb{Z}\}$. Suppose $I \neq \emptyset$. If we decrease each u_i, $i \in I$, by ε and increase each v_j, $j \in J$, by ε, then the right-hand side of (87) is linear in ε; suppose it non-increases with ε. Let $\varepsilon = \min_{i \in I}(u_i - \lfloor u_i \rfloor)$. We obtain at least one more integer among the u's, so to obtain a contradiction it is enough to check that (88)–(91) still hold. The restriction (88), for example, may cause us a problem only if $0 \in I \setminus J$. Then $v_0 \in \mathbb{Z}$ and, by definition, ε is at most the fractional part of $u_0 + v_0$ which would suffer the decrement by ε without crossing the integer 1. The claim is proved.

We may assume that $v_0 = 0$, because we can add some integer ε to each u_i and subtract ε from each v_i, $i \in [0,n]$, without affecting (87)–(91). Also, we can make $u_0 = 1$, because we can subtract $\varepsilon > 0$ from u_0 and add ε to each v_i, $i \in [n]$. Hence, $k = n + \min(\sum_{i=1}^n u_i + \sum_{j=1}^n v_j)$, given conditions $u_i \geq 0$, $v_i \geq -1$, $i \in [n]$, and (91). It is easy to see that in our (integral) solution, each u_i is either 0 or 1 and each v_i is either -1 or 0. Let $X = \{j \in [n] : v_j = -1\}$ and let $X' = \{i \in [n] : \exists j \in X \ (i,j) \in E(D)\}$, that is, X' consists of vertices sending at least one edge to X.

To satisfy (91) we must have $u_i = 1$ for each $i \in X'$. Also, if we set $u_i = 0$ for $i \in [n] \setminus X'$, then (88)–(91) are still satisfied while the linear function in (87) does not increase. Hence, we may assume that $X' = \{i \in [n] : u_i = 1\}$; then $n - k = |X| - |X'|$.

Let $A = X \setminus X'$ and $B = X' \setminus X$. Let $P = \{x_1, \ldots, x_l\}$ be a directed path in D with $x_1, x_l \in A$, $l \geq 2$. As $x_1 \notin X'$, we conclude $x_2 \notin X$. As $x_l \in X$, there must be $i \in [2, l-1]$ such that $x_i \notin X$ but $x_{i+1} \in X$. By definition, $x_i \in B$. Similarly, any cycle intersecting A intersects B. By (86) we obtain,

$$m(D) \leq n - k = |X| - |X'| = |A| - |B| \leq M(D),$$

which was required. ∎

Remark. Graham Brightwell considerably simplified our proof shortly after it had been announced. Let us outline his argument which exploits Hall's theorem.

Given a digraph D, consider the bipartite graph G on two copies of $V(D)$, say $X = \{v^\vee : v \in V(D)\}$ and $Y = \{v^\wedge : v \in V(D)\}$, where we connect $u^\vee$ to $v^\wedge$ if and only if $(u,v) \in E(D)$. It is easy to check that the number of edges missing in a maximum matching in G equals $m(D)$. By a version of Hall's theorem, this number equals the maximum of $|Z| - |\Gamma(Z)|$ over $Z \subset X$. Choose any extremal set Z. Now, it is routine to check that

$$\begin{aligned} A &= \{v \in V(D) : v^\vee \in Z,\ v^\wedge \notin \Gamma(Z)\}, \\ B &= \{v \in V(D) : v^\vee \notin Z,\ v^\wedge \in \Gamma(Z)\}, \end{aligned}$$

are two sets exhibiting $m(D) = |A| - |B| \leq M(D)$.

The following corollary is obtained by applying Theorem 64 to the Hasse diagram of $\mathcal{P}$.

Corollary 65 *For any poset $\mathcal{P}$, the minimal number m of skipless chains partitioning it equals the maximal value of $|A| - |B|$ over all disjoint sets $A, B \subset \mathcal{P}$ such that any skipless chain containing two elements from A intersects B.* ▮

Of course, the minimal size m of an edge decomposition of $\mathcal{P}$ can be computed by applying Corollary 65 to $L(\mathcal{P})$. However, we present a direct proof which is short and gives a direct algorithm for constructing such a partition. It turns out that to compute $M(L(\mathcal{P}))$ it is enough to consider only pairs $A, B \subset L(\mathcal{P})$ of the following rather special form: take a partition $X \cup Y = \mathcal{P}$ and let $A = \{(x \lessdot y) \in L(\mathcal{P}) : x \in X, y \in Y\}$ and $B = \{(y \lessdot x) \in L(\mathcal{P}) : x \in X, y \in Y\}$.

We state the result in terms of digraphs. Let $e(X,Y)$, $X, Y \subset V(D)$, denote the number of the edges in D starting in X and ending in Y and $M(X,Y) = e(X,Y) - e(Y,X)$.

Theorem 66 *The minimal number m of directed paths partitioning the edge set of a cycle-free digraph D is equal to*

$$M(D) = \max\{M(X,Y) : X \cup Y = V(D),\ X \cap Y = \emptyset\}.$$

Proof. It is immediate that $m \geq M$ because for any partition $X \cup Y = V(D)$ and any path P the removal of the edges on P can decrease $M(X,Y)$ by at most one. To prove the reverse inequality by induction on $|E(D)|$ it is enough to show that, for the graph D' obtained from D by the removal of the edges of a maximal path $P = (x_1, \ldots, x_k)$, we have $M(D') < M(D)$.

To show this take a partition $X \cup Y = V(D')$ with

$$M(D') = M(X,Y) = e(X,Y) - e(Y,X).$$

Since P is maximal and D is acyclic there is no $y \in V(D)$ with $(y, x_1) \in E(D)$. Therefore if $x_1 \in Y$ we can move x_1 to X without decreasing $M(X,Y)$. Likewise we may assume $x_k \in Y$. But if we add back the edges of P we will increase $M(X,Y)$ by 1: if moving along P we change side from Y to X i times, then we go from X to Y $i+1$ times. This shows that $M(D') < M(D)$ as required. ▮

Remark. Incidentally, we discovered an algorithm producing an optimal edge decomposition: select and remove maximal paths one by one.

15 Symmetric Edge Partitions of Cubes

The result of de Brujin, Tengbergen and Kruyswijk [BTK51] (see [Bol86, Theorem 4.1] or [And87, Section 3.1] for a proof) asserts that $\mathcal{B}_n = (2^{[n]}, \subset)$ is a *symmetric chain order*, that is, admits a decomposition into *symmetric chains*. (A chain $x_1 < \ldots < x_k$ in a ranked poset $(\mathcal{P}, r)$ is called *symmetric* if it is skipless and $r(x_1) = r(\mathcal{P}) - r(x_k)$.) This was strengthened by Anderson [And67] and Griggs [Gri77], who showed that a LYM poset $\mathcal{P}$ with a unimodal symmetric rank-sequence is a symmetric chain order. (Note that the number of chains is $w(\mathcal{P})$—minimal possible.)

The latter result is applicable to $L(\mathcal{B}_n)$ which, as a regular poset, has the LYM property, see e.g. [Eng97, Corollary 4.5.2]. However, this way we obtain a purely existential result while one would wish to have an explicit decomposition. Here we provide an explicit construction, which like that of Leeb (unpublished) and Greene and Kleitman [GK76] on $\mathcal{B}_n$, utilizes bracket representations.

Theorem 67 *$L(\mathcal{B}_n)$ is a symmetric chain order. In other words, $\mathcal{B}_n$ admits an edge decomposition into symmetric chains.*

Proof. Assume that the numbers $1, \ldots, n$ are placed on a circle clockwise in this order. Let σ denotes the *shift* which maps every element to the next position clockwise: $\sigma(k) = k + 1 \pmod{n}$ and let $\sigma^{(i)}$ be its ith iterate. (These are also referred to as *rotations*.) For the clarity of language we use the same symbol σ for the corresponding action on the vertex set and the edge set of Q_n. We will produce a σ-invariant edge partition.

We build, inductively on n, a family $\mathcal{F}_n$ of n-element sequences, starting for the case $n = 1$ with the family $\mathcal{F}_1 = \{\ (\ \}$. To build $\mathcal{F}_{n+1}$ apply Operations A and B to every sequence $F \in \mathcal{F}_n$ and let $\mathcal{F}_{n+1}$ comprise the resulting sequences. Operation A: add '(' to the right of F. Operation B: add ')' to the right of F and throw away the resulting sequence if it does not contain free elements (i.e. if all its parentheses can be paired).

Proceeding in this way we obtain, for example,

$$\begin{aligned}\mathcal{F}_2 &= \{\,((\,\},\\ \mathcal{F}_3 &= \{\,(((,\ (()\,\}\\ \mathcal{F}_4 &= \{\,((((,\ ((()),\ (()(\,\}.\end{aligned}$$

It is easy to see that $\mathcal{F}_n$ is the set of all n-sequences beginning with '(' which is a free element. (In particular, all right parentheses are paired.)

For any sequence $F \in \mathcal{F}_n$ we build the corresponding chain C_F in $L(\mathcal{B}_n)$ which has length t, where t is the number of free members of F. To obtain the $(*)$-description of the ith element of C_F, $i \in [t]$, we reverse in F the last $i-1$ free parentheses and replace the ith free element (when counted from the right) by star $*$. Thus, for example, '$(()(()$' gives $(()*()$ and $*()))()$ which correspond to the following chain in $L(\mathcal{B}_6)$:

$$(\{3,6\} \sqsubset \{3,4,6\}) \lessdot (\{3,4,6\} \sqsubset \{1,3,4,6\}).$$

It is easy to see that every C_F is a symmetric chain. We claim that

$$\mathcal{D}_n = \{\sigma^{(j)}(C_F) : F \in \mathcal{F}_n,\ j = 0,\ldots,n-1\}$$

is the required edge partition.

We have to prove that for every element $x = (A \sqsubset B)$ in $L(\mathcal{B}_n)$ there are unique $F \in \mathcal{F}_n$ and $j \in [0, n-1]$ such that $x \in \sigma^{(j)}(C_F)$. First we show how to find at least one such pair (F, j).

Step 1. Write x in the $(*)$-representation. **Step 2.** Rotate the pattern to bring the star to position 1 and then identify all free parentheses. Clearly, if disregarding the paired elements, our sequence is '$*)\ldots)(\ldots($.' **Step 3.** Rotate again so that the first free left parenthesis identified in Step 2 (or the star itself if no '(' is free) is moved to position 1. Let j be the number of positions that the star was moved anticlockwise by Steps 2 and 3 combined. **Step 4.** Replace the star and all free right parentheses identified in Step 2 by left parentheses. Let F be the resulting sequence.

Obviously, when we pair brackets in Step 4, we obtain the same sets of free/paired elements as in Step 2. This implies that $F \in \mathcal{F}_n$ as it starts with free '(' and that $x \in C_F$ as required. Here is an illustration for $x =$

$(\{1,6,7\} \sqsubset \{1,4,6,7\}) \in L(\mathcal{B}_8)$:

$$\begin{array}{llllllllll}
\text{Step 1:} &) & (& (& * & (&) &) & (& \\
\text{Step 2:} & * & \boxed{(\,)} &) & \boxed{(\,)} & (& (& & & \\
\text{Step 3:} & (& (& * & \boxed{(\,)} &) & \boxed{(\,)} & & & (\text{and } j=1) \\
\text{Step 4:} & (& (& (& (&) & (& (&) & (\text{this is } F)
\end{array}$$

The uniqueness of (F, j) can be established in different ways. One, which actually gives an alternative definition of $\mathcal{D}_n$, is the following. Given the $(*)$-representation of x, for $0 \leq i \leq n-1$ let $g(i) = l_i - r_i$, where l_i and r_i are respectively the number of left and right parentheses in the i positions preceding '$*$' clockwise. If $x \in \sigma^j(C_F)$ then j is the smallest element in $[0, n-1]$ on which g achieves its maximum. Why? Just pair the brackets in the $(*)$-representation of $\sigma^{-j}(x) \in C_F$, e.g.

$$(\;\boxed{(\,(\,)\,(\,)\,)}\;\boxed{(\,)}\;(\,*\,)\;\boxed{(\,)}\;)\;\boxed{(\,)}\;,$$

and notice that any paired block (boxed regions) contributes 0 to g while any right-hand-sided part of it contributes a strictly negative value. Now, the maximum of g is the number of free left parentheses and this is achieved for first time when considering the segment preceding the star, as required.

But now, once that j has been identified, there trivially could not be two suitable F's. ∎

For the remainder of this part let $\mathcal{D}_n$ denote the edge decomposition of $\mathcal{B}_n$ constructed above. It has the following properties.

Theorem 68 *Let $C = (A_1 \sqsubset \ldots \sqsubset A_k)$ be one of the chains in $\mathcal{D}_n$. If $A_{i+1} = A_i \cup \{a_i\}$, then the elements $a_1, \ldots, a_{k-1}$ are situated on the circle in this other anticlockwise and between a_i and a_{i+1} (anticlockwise) there is an even number of places. For each $i \in [k-3]$, there is an element $(B \sqsubset B')$ belonging to a chain of $\mathcal{D}_n$ shorter than C such that*

$$A_i \sqsubset B \sqsubset B' \sqsubset A_{i+3}. \tag{92}$$

Proof. Take the sequence $F \in \mathcal{F}_n$ giving rise to C. (We may assume $j = 0$.) The fact that in F every pair of consecutive free elements contains only paired brackets in between implies the firt part of the theorem.

To show the second claim, let F' be the sequence F with the $(i+1)$st free left bracket (if counted from the right) replaced by ')' which is then paired with the $(i+2)$nd free element:

$$
\begin{array}{lcccccccccccccl}
F: & (& \square & (& \square & (& \square & (& \square & (& \square & (& \square & & \\
 & & & & & & & & & * & \multicolumn{4}{c}{\ldots\ldots\ldots\ldots} & A_i \\
 & & & * & \multicolumn{10}{c}{\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots} & A_{i+3} \\
F': & (& \square & (& \square & \multicolumn{2}{c}{\boxed{(\square)}} & & \square & (& \square & (& \square & & \\
 & & & & & & & & & * & \multicolumn{4}{c}{\ldots\ldots\ldots\ldots} & B \\
 & & & * & \multicolumn{10}{c}{\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots} & B'
\end{array}
$$

The new sequence corresponds to a chain of length $k-2$ and its ith and $i+1$st elements obviously satisfy the required property. ∎

We define the *complementary chain* $\overline{C}$ of a chain C by replacing every element by its complement, ie. if $C = (A_1 \subset \ldots \subset A_k)$ then $\overline{C} = (\overline{A_k} \subset \ldots \subset \overline{A_1})$.

Lemma 69 *Two elements $x_1 = (A_1 \sqsubset B_1)$ and $x_2 = (A_2 \sqsubset B_2)$ of $L(\mathcal{B}_n)$ can belong simultaneously to $\mathcal{D}_n$ and $\overline{\mathcal{D}_n}$ only if $n = 2k$ is even and $\{|B_1|, |B_2|\} = \{k, k+1\}$.*

Proof. Let $i_h \in [n]$ be the element of B_h not in A_h, $h = 1, 2$, and let pairs (F, j) and (F', j') give rise to chains $C, C' \in \mathcal{D}_n$ such that C' contains x_1 and x_2 while C contains $\overline{x_1}$ and $\overline{x_2}$ respectively. Assume that $j' = 0$ and $x_1 < x_2$, ie. $B_1 \subset A_2$.

In F' i_2 precedes i_1 and we claim that F' does not contain a free element between them. Indeed, if it be in the position $y \in [n]$ then $y \in \overline{B_1}$ and $y \notin \overline{A_2}$, that is, $\sigma^{-j}(y)$ must be a free element in F. But then $\sigma^{-j}(y)$ must lie between $\sigma^{-j}(i_1) < \sigma^{-j}(i_2)$. (In C the element $\overline{x_2}$ comes before $\overline{x_1}$.) This contradiction (on one hand the elements i_2, y, i_1 go clockwise, on the other—anticlockwise) proves the claim.

Thus all the elements between i_2 and i_1 are paired in F'; therefore $B_1 = A_2$ and there must be the same number of left and right parentheses in this interval. Considering $\overline{x_2}, \overline{x_1} \in C$ we show the analogeous statement about the elements between i_1 and i_2 (if going clockwise), which clearly implies the claim. ∎

16 Applications of the Partition $\mathcal{D}_n$

We would like to include here some applications of the edge partition $\mathcal{D}_n$ built in Theorem 67. Basically, we are inspired by known results where a symmetric vertex decomposition of $\mathcal{B}_n$ is used. We refer the reader to Section 3.4 of Anderson's book [And87] for an exposition of a few results of this type. I am grateful to Ian Anderson for drawing my attention to some other applications not surveyed in his book.

16.1 On the Number of Antichains in $L(\mathcal{B}_n)$

Let us consider the following question: what is $\varphi(L(\mathcal{B}_n))$, the number of antichains in $L(\mathcal{B}_n)$? The computation of $\varphi(\mathcal{B}_n)$ is an old and difficult problem; a complicated asymptotic formula was established by Korshunov [Kor81].

Here we provide some rough estimates of $\varphi(L(\mathcal{B}_n))$ by applying ideas of Hansel [Han66] who showed that $2^N \le \varphi(\mathcal{B}_n) \le 3^N$, where $N = w(\mathcal{B}_n) = \binom{n}{\lfloor n/2 \rfloor}$.

Considering all possible subsets of the largest antichain of $L(\mathcal{B}_n)$ we obtain trivially $\varphi(L(\mathcal{B}_n)) \ge 2^m$, where $m = w(L(\mathcal{B}_n)) = \lceil n/2 \rceil \binom{n}{\lfloor n/2 \rfloor}$.

On the other hand, observe that an antichain $A \subset L(\mathcal{B}_n)$ is uniquely determined by the ideal $\Delta(A) = \{x \in L(\mathcal{B}_n) : \exists\ a \in A\ x \le a\}$. Consider any $C = (x_1 \lessdot \ldots \lessdot x_l) \in \mathcal{D}_n$. By Theorem 68 for $3 \le i \le l-2$ we can find y_i in a shorter chain with $x_{i-2} < y_i < x_{i+2}$. Knowing $\Delta(A) \cap C'$ for every $C' \in \mathcal{D}_n$ shorter than C we know $\Delta(A) \cap \{y_3, \ldots, y_{l-2}\}$. But then it is easy to check that only for at most 4 elements of C we are unable to deduce whether it is in $\Delta(A)$, and therefore $\Delta(A) \cap C$ can assume at most 5 possible values. Considering consecutively the chains of $\mathcal{D}_n$ in some size-increasing order we conclude that $\varphi(L(\mathcal{B}_n)) \le 5^m$.

16.2 Orthogonal Partitions of $L(\mathcal{B}_n)$

Two chains in a poset $\mathcal{P}$ are called *orthogonal* if they have at most one common element. Two vertex chain partitions $\mathcal{D}$ and $\mathcal{D}'$ are *orthogonal* if any $C \in \mathcal{D}$ is orthogonal to any $C' \in \mathcal{D}'$. A result of Shearer and Kleitman [KS79] (see [And87, Section 3.4]) asserts that there exist two orthogonal chain decompositions of $\mathcal{B}_n$ into $\binom{n}{\lfloor n/2 \rfloor}$ chains each.

What can be said about $L(\mathcal{B}_n)$? If n is odd, then $\mathcal{D}_n$ and $\overline{\mathcal{D}_n}$ are orthogonal by Lemma 69, where $\mathcal{D}_n$ is the decomposition built in Theorem 67.

Theorem 70 *For odd n there is a pair of orthogonal symmetric chain decompositions of $L(\mathcal{B}_n)$.* ∎

Remark. Unfortunately, we do not know if the corresponding claim is true for even n.

The result of Baumert, McEliece, Rodermich and Rumsey [BMRR80] (for a proof, see [And87, Section 3.4.3] or [Bol86, Section 6]) states that posets admitting a pair of orthogonal decompositions satisfy the probabilistic form of Sperner's theorem, which in our case, by Theorem 70, is the following.

Corollary 71 *If two elements x and y in $L(\mathcal{B}_n)$, odd n, are chosen independently with arbitrary probability distribution (same for both elements) then $P\{x \leq y\} \geq w(L(\mathcal{B}_n))^{-1}$.* ∎

16.3 A Storage and Retrieval Problem

Suppose we maintain a database with n records which we number from 1 to n and we wish to organize an efficient searching. We assume that we have queries $Q_1, \ldots, Q_M$ each of which we identify with the set of records satisfying it, that is, $Q_i \subset [n]$ and these subsets are not necessarily distinct. One idea, see Ghosh [Gho75], is to find a sequence X of elements of $[n]$ such that every Q_i occurs in X as a subsequence of consecutive terms so that every Q_i can be defined by a starting position in X and the size of Q_i.

In connection with this Lipski [Lip78] considered the following problem. Find the shortest sequence of elements of $X = [n]$ such that X contains every $A \subset [n]$ as a subsequence of $|A|$ consecutive terms. He showed that s_n, the length of an optimal sequence, satisfies

$$\left(\frac{2}{\pi n}\right)^{1/2} 2^n \leq (1 + o(1)) s_n \leq \left(\frac{2}{\pi}\right) 2^n. \tag{93}$$

As far as I know, these might be the best known bounds to date.

Here we consider a similar problem. Namely, we ask what is the value of t_n, the shortest length of a sequence X such that for every $A \sqsubset B \subset [n]$ the sequence X contains A as a subsequence of $|A|$ consecutive terms preceded by x, where $\{x\} = B \setminus A$. Such a situation can happen if every query is a set with a selected point. For example, we search in a dictionary, the allowed queries are of the form "Find *word*" and the answer should give the entry where *word* is defined plus all relevant entries. Applying the ideas of Lipski [Lip78] we find the following upper and lower bounds.

Theorem 72

$$\left(\frac{n}{2\pi}\right)^{1/2} 2^n \leq (1+o(1))t_n \leq \left(\frac{n}{\pi}\right) 2^n. \tag{94}$$

Proof. As the number of different pairs $A \sqsubset A \cup \{x\}$ with $|A| = \lfloor n/2 \rfloor$ which can lie within a sequence of length m does not exceed $m - \lfloor n/2 \rfloor$ we conclude that

$$t_n \geq \lceil n/2 \rceil \binom{n}{\lfloor n/2 \rfloor} + \lfloor n/2 \rfloor$$

which implies the lower bound in (94) by Stirling's formula.

On the other hand, associate with every chain $C = (A_1 \sqsubset \ldots \sqsubset A_q)$ in $\mathcal{D}_n$ a sequence of elements of $[n]$ which contains first the elements of A_1 in any order which then are followed by $a_2, \ldots, a_q$, where $\{a_i\} = A_i \setminus A_{i-1}$, $i = 2, \ldots, q$.

Let $[n] = S \cup T$ be a partition of $[n]$ into 2 parts of (nearly) equal sizes. Let $\phi_1, \ldots \phi_k$ be the sequences corresponding to a symmetric vertex decomposition of 2^T. Also, let $\psi_1, \ldots, \psi_l$ be the sequences corresponding to a symmetric edge decomposition of 2^S, each sequence being reversed.

Clearly, for every $A \subset S$ there exists ϕ_i containing A as the first consecutive $|A|$ terms and for every $A \sqsubset A \cup \{x\} \subset T$ there exists ψ_j containing, at the end, A preceded by x.

Now consider the sequence

$$X = \psi_1\phi_1\psi_1\phi_2 \ldots \psi_1\phi_k\psi_2\phi_1\psi_2\phi_2 \ldots \psi_l\phi_k.$$

Take any $A \subset [n]$ and $x \in T \setminus A$. There is ψ_i containing x at the end followed by $A \cap T$ and ϕ_j containing $A \cap S$ as an initial subsequence. Therefore, X_1 contains x followed by A. Interchanging S and T, we write a sequence X_2 containing every pair $A \sqsubset A \cup \{x\}$ with $x \in S$. The sequence $X = X_1X_2$ is the required (and explicitly constructed) sequence. It is easy to see that the average size of a sequence corresponding to a chain of a symmetric vertex or edge decomposition of $\mathcal{B}_n$ is $(\frac{1}{2}+o(1))n$. Therefore,

$$t_n \leq |X| \leq 2\left(\tfrac{1}{2}+o(1)\right) nkl$$

which gives the desired upper bound by Stirling's formula. ∎

16.4 One Numerical Problem

There exists a so called Audit Expert Mechanism which can be used to protect small statistical databases, see Chin and Ozsoyoglu [CO82]. To find an optimal mechanism the following problem has to be solved. Suppose we operate with n-tuples of non-zero real numbers $a_1, \ldots, a_n$ and we want to find what is the maximum possible number of subsets $I \subset [n]$ such that a_I is equal to either 0 or 1. (Here and later we denote $a_I = \sum_{i \in I} a_i$.) The best possible bound of $\binom{n+1}{\lfloor (n+1)/2 \rfloor}$ was found by Miller, Roberts and Simpson [MRS91] and all extremal sequences were characterized by Miller and Sarvate [MS95]. Both papers make use of the existence of a symmetric chain decomposition of $\mathcal{B}_n$.

Here, applying a symmetric chain decomposition of $L(\mathcal{B}_n)$, we can find K, the maximal possible number of elements $(I \sqsubset J) \in L(\mathcal{B}_n)$ such that $\{a_I, a_J\} = \{0,1\}$, over all real sequences $a_1, \ldots, a_n$. Actually, we can allow zero entries for, as we will see later, this does not affect K. Apparently, this problem does not have such an application like that of the original problem, but it might be of some interest especially as an unexpected application of a symmetric chain decomposition of $L(\mathcal{B}_n)$.

The expression $(a)^i$ is a shorthand for a repeated i times. Also we assume that all n-tuples have their entries ordered non-decreasingly.

Theorem 73 *For $n \geq 2$ we have*

$$K = \lceil n/2 \rceil \binom{n}{\lfloor n/2 \rfloor}, \tag{95}$$

and this value is achieved for and only for the following sequences. For $n = 2k$, $((-1)^k, (+1)^k)$, $((-1)^{k-1}, (+1)^{k+1})$ and $((-1)^{k-1}, 0, (+1)^k)$. For $n = 2k+1$, $((-1)^k, (+1)^{k+1})$.

Proof. Let m be the largest index for which $a_m < 0$. Define $f : 2^{[n]} \to 2^{[n]}$ by the formula

$$f(I) = I \triangle [m] = (I \setminus [m]) \cup ([m] \setminus I), \qquad I \subset [n].$$

One can easily check that $I \subset J \subset [n]$ implies $a_{f(I)} \leq a_{f(J)}$.

$\mathcal{D}_n$ can be viewed as a collection of symmetric chains in $2^{[n]}$. Let $X_r \sqsubset \ldots \sqsubset X_{n-r}$ be one such chain. The sequence

$$a_{f(X_r)}, \ldots, a_{f(X_{n-r})}$$

is non-decreasing and therefore 0 and 1 can occur side by side there at most once. As every $A \sqsubset B$ is present in exactly one chain and f is a bijection preserving or reversing the $\sqsubset$-relation, K does not exceed the total number of chains, which gives the required upper bound.

A moment's thought reveals that a necessary and sufficient condition for an n-tuple to be optimal is the following. If $n = 2k+1$ then for every $A \sqsubset B \subset X$, $|A| = k$, we have $a_{f(A)} = 0$ and $a_{f(B)} = 1$. If $n = 2k$ then for every $A \sqsubset B \sqsubset C \subset X$, $|A| = k-1$, among the numbers

$$a_{f(A)} \leq a_{f(B)} \leq a_{f(C)} \tag{96}$$

there is a 0 adjacent a 1.

This condition is fulfilled for the sequences mentioned in the statement. Indeed, let us consider $((-1)^k, (+1)^{k+1})$, for example. Here $m = k$ and for any $A \sqsubset B$ with $|A| = k$ we have

$$a_{f(A)} = a_{A \triangle [k]} = (-1)(k-s) + (k-s) = 0, \tag{97}$$

where $s = |A \cap [k]|$. Similarly, $a_{f(B)} = 1$ so the sequence is optimal.

We claim that these are essentially the cases of the equality. Let us do the harder case $n = 2k$. If, for some $i \neq j$, we have $a_i \neq \pm 1$ and $a_j \neq \pm 1$, then $A \sqsubset A \cup \{i\} \sqsubset A \cup \{i,j\}$ with any $A \in X^{(k-1)}$, $A \not\ni i, j$, obviously violates the condition. If, for exactly one i, we have $a_i \neq \pm 1$, then considering $A \sqsubset A \cup \{i\} \sqsubset C$ we conclude that $a_{f(A \cup \{i\})} = 0$ for any $A \in (X \setminus \{i\})^{(k-1)}$. Suppose $a_i \geq 0$, for example. Then $a_{f(A \cup \{i\})} = k - j - 1 + a_i = 0$, where j is the total number of elements equal to -1 (so $2k - 1 - j$ elements equal $+1$). If

$a_i = 0$, then we have the third example mentioned in the theorem. If $a_i \geq 2$ then $j \geq k+1$ and any sequence (96) with $C \not\ni i$ violates the condition. Finally, if $|a_i| = 1$ for every i then arguing as in (97) we deduce that we can have either k or $k+1$ positive entries. ∎

17 Characterization of Line Posets

For graphs we know that we can characterize line graphs in terms of forbidden induced subgraphs (Beineke [Bei68]) and we can reconstruct a connected graph G given $L(G)$ except for $L(G) = K_3$ when G is either K_3 or $K_{1,3}$.

Here we ask ourselves when a given poset $\mathcal{L}$ is the line poset of some $\mathcal{P}$ and what information about $\mathcal{P}$ can be reconstructed from $L(\mathcal{P})$. (Of course, it is implicitly understood that we operate with isomorphism classes of posets.) While for line graphs there are nine forbibben configuration, for line posets we have only two (or infinitely many, depending on how we look at it).

Note that $L(\mathcal{P})$ cannot contain elements w, x, y, z such that $w \lessdot y$, $x \lessdot y$, $w \lessdot z$ but $x \not\lessdot z$; call this configuration N. Indeed, if y and z cover w they must be of the form $(a \lessdot b)$, $(a \lessdot c)$, where $w = (d \lessdot a)$, some $a, b, c, d \in \mathcal{P}$. Then the relation $x \lessdot y$ implies that $x = (e \lessdot a)$ which implies that $x \lessdot z$.

Also, $L(\mathcal{P})$ cannot contain the configuration C_n, $n \geq 3$, made of elements y and $x_1, \ldots, x_n$ such that $x_1 \lessdot y \lessdot x_n$ and $x_i \lessdot x_{i+1}$, for $i \in [n-1]$. Indeed, suppose the contrary. Clearly, $\mathcal{P}$ contains elements $z_0 \lessdot z_1 \lessdot \ldots \lessdot z_n$ such that $x_i = (z_{i-1} \lessdot z_i)$. But y covers the same element as x_2 and is covered by the same element as x_{n-1}, so $y = (z_1 \lessdot z_{n-1})$ and $n = 3$; but then $y = x_2$, which is a contradiction.

For a poset $\mathcal{P}$ let $T(\mathcal{P}) = (\mathcal{C}, k, l, u)$ be the quadruple with $\mathcal{C}$ being a subposet of $\mathcal{P}$ spanned by the non-extremal elements, that is by $\{a \in \mathcal{P} : \exists\, b, c \in \mathcal{P},\ b < a < c\}$ and k is the number of pairs $(a \lessdot b)$ with $a, b \in \mathcal{P} \setminus \mathcal{C}$ while the functions $l, u : \mathcal{C} \to \mathbf{N_0}$ are given by

$$\begin{aligned} l(a) &= |\{x \in \mathcal{P} \setminus \mathcal{C} : x \lessdot a\}|, \\ u(a) &= |\{x \in \mathcal{P} \setminus \mathcal{C} : x \gtrdot a\}|, \qquad a \in \mathcal{P}. \end{aligned}$$

It is easy to see that $T(\mathcal{P})$ determines $L(\mathcal{P})$ uniquely.

The following theorem states that the above examples provide a complete answer to our two questions.

Theorem 74 *A poset $\mathcal{L}$ is isomorphic to $L(\mathcal{P})$ for some $\mathcal{P}$ if and only if $\mathcal{L}$ contains neither configuration N nor any of C_n, $n \geq 3$. Furthermore, $T(\mathcal{P})$ determines $L(\mathcal{P})$ and can be reconstructed from it.*

Proof. Given a poset $\mathcal{L}$ without N or C_n let X be two disjoint copies of its vertex set, namely $X = \{x^\wedge, x^\vee : x \in \mathcal{L}\}$. Let $x^\wedge \sim y^\vee$ if $x \lessdot y$; let $x^\wedge \sim y^\wedge$ if, for some $s \in \mathcal{L}$, we have $s \gtrdot x$ and $s \gtrdot y$; let $x^\vee \sim y^\vee$ if, for some $s \in \mathcal{L}$, $s \lessdot x$ and $s \lessdot y$.

We claim that $\sim$ is an equivalence relation. Indeed, if $x^\wedge \sim y^\wedge$ and $y^\wedge \sim z^\wedge$ then there are $s, t \in \mathcal{L}$ such that $x, y \lessdot s$ and $y, z \lessdot t$. But then t must cover x for otherwise x, y, s, t would span a forbidden configuration. So $x, z \lessdot t$ and $x^\wedge \sim z^\wedge$. The remaining cases are equally easy.

Let $\overline{x}$ denote the equivalence class of $x \in X$. Define the poset $\mathcal{P}$ (also denoted by $L^{-1}(\mathcal{L})$) on $V = X/{\sim} = \{\overline{x} : x \in X\}$ by $A < B$, $A, B \in V$ iff in $\mathcal{L}$ there exist $y \leq z$ with $y^\vee \in A$ and $z^\wedge \in B$. One can check that this is indeed an ordering. For example, to check its transitivity, let $A < B$ and $B < C$, choose $w \leq x$ and $y \leq z$ in $\mathcal{L}$ with $w^\vee \in A$, $x^\wedge, y^\vee \in B$ and $z^\wedge \in C$; then $x^\wedge \sim y^\vee$ implies that $w \leq x \lessdot y \leq z$ and $A < C$.

Let us show that $\overline{x^\wedge}$ covers $\overline{x^\vee}$. Assuming the contrary we find $z \geq y$ and $w \geq v$ in $\mathcal{L}$ with $z^\wedge \sim x^\wedge$, $y^\vee \sim w^\wedge$ and $v^\vee \sim x^\vee$. By the definition of $\sim$, some $t \in \mathcal{L}$ covers both x and z, some $s \in \mathcal{L}$ is covered by both x and v and $v \leq w \lessdot y \leq z$—which implies that $\mathcal{L}$ contains some C_n, which is forbidden.

We claim that $\mathcal{L} \cong L(\mathcal{P})$ via the map F which sends $x \in \mathcal{L}$ to $(\overline{x^\vee} \lessdot \overline{x^\wedge})$. First note that F is an order preserving map: if $x \gtrdot y$ in $\mathcal{L}$ then $\overline{x^\vee} \sim \overline{y^\wedge}$ which implies $F(x) \gtrdot F(y)$ as desired. Next, F is injective for if $F(x) = F(y)$ then $x^\wedge \sim y^\wedge$ and $x^\vee \sim y^\vee$ which implies that for some w and z we have $w \lessdot x \lessdot z$ and $w \lessdot y \lessdot z$; but as $\mathcal{L}$ does not contain configuration C_3 we conclude that $x = y$. To show that F is surjective take any $(A \lessdot B) \in L(\mathcal{P})$. As $A < B$, for some $\mathcal{L}$-elements $x \leq y$ we have $A = \overline{x^\vee}$, $B = \overline{y^\wedge}$. But it is easy to see that $\overline{x^\wedge} \leq \overline{y^\wedge}$, which implies $(A \lessdot B)$ equals $(\overline{x^\vee} \lessdot \overline{x^\wedge}) = F(x)$. Finally, if $F(x) \lessdot F(y)$ then $x^\wedge \sim y^\vee$ and $x \lessdot y$. This proves completely that $\mathcal{L} \cong L(\mathcal{P})$.

In the second part it is enough to show that for any poset $\mathcal{R}$ we have $T(\mathcal{R}) \cong T(\mathcal{P})$, where $\mathcal{P} = L^{-1}(\mathcal{L})$, $\mathcal{L} = L(\mathcal{R})$. To build a natural isomorphism $H : \mathcal{C}(\mathcal{R}) \to \mathcal{C}(\mathcal{P})$ take, for any element $a \in \mathcal{C}(\mathcal{R})$, some $b \lessdot a$ which exists as a is a non-extremal element of $\mathcal{R}$. Now let $H(a) = \overline{x^\wedge}$, where $x = (b \lessdot a) \in \mathcal{L}$ and $\sim$ is as above. To show that H is well defined, let b' be another choice of b and denote $y = (b' \lessdot a)$. Let c be an element covering a. Then $(a \lessdot c)$ covers in $\mathcal{L}$ both x and y, so by the definition of $\mathcal{P}$ we have $x^\wedge \sim y^\wedge$. Also, $H(a) \in \mathcal{P}$ is not extremal as

$$\overline{(b \lessdot a)^\vee} < H(a) < \overline{(a \lessdot c)^\wedge}.$$

Next, H is an order-preserving bijection. Indeed, let $a \gtrdot b$ in $\mathcal{C}(\mathcal{R})$. Choose $c \lessdot b$. Then $H(a) = \overline{(b \lessdot a)^\wedge}$ and $H(b) = \overline{(c \lessdot b)^\wedge}$. But $(c \lessdot b)^\wedge \sim (b \lessdot a)^\vee$ and we have $H(a) > H(b)$ by the definition of the order on $\mathcal{P}$. To show that H is injective choose any $a, a' \in \mathcal{C}(\mathcal{R})$. Then $H(a) = H(a')$ implies that $y = (c \lessdot a)^\wedge \sim y' = (c' \lessdot a')^\wedge$, some $c, c' \in \mathcal{R}$. Therefore there is $x \in \mathcal{L}$ covering both y and y' which implies $a = a'$ in $\mathcal{R}$ as required. To establish the surjectivity of H consider $x = \overline{(a \lessdot b)^\vee} \in \mathcal{C}(\mathcal{P})$, for example. Observe first that $a \in \mathcal{R}$ is not extremal. Indeed, take any $y \in \mathcal{P}$ covered by x; as we have already shown any pair $y \lessdot x$ is of the form $\overline{(c \lessdot d)^\vee} \lessdot \overline{(c \lessdot d)^\wedge}$ which implies $d = a$ and $c \lessdot a$. Now $H(a) = \overline{(c \lessdot a)^\wedge} = x$ as required. Again, any two adjacent elements of $\mathcal{C}(\mathcal{P})$ can be represented as $\overline{(a \lessdot b)^\vee} \lessdot \overline{(a \lessdot b)^\wedge}$ and then they are the images of two adjacent elements, $a \lessdot b$ of $\mathcal{C}(\mathcal{R})$, which implies that $\mathcal{C}(\mathcal{P}) \cong \mathcal{C}(\mathcal{R})$.

Finally, as $\mathcal{P}$ and $\mathcal{R}$ give rise to naturally isomorphic line posets, in the sense that

$$F(a \lessdot b) = \left(\overline{(a \lessdot b)^\vee} \lessdot \overline{(a \lessdot b)^\wedge}\right) = (H(a) \lessdot H(b)), \quad a, b \in \mathcal{C}(\mathcal{R}),$$

our mapping H preserves k, l and u, which are naturally reconstructible from the line poset. ∎

Part IV

Enumeration Results for Trees

18 Introduction

The notion of a tree and its different extensions to hypergraphs play an important role in discrete mathematics and computer science. We will dwell upon the following, rather general, definition suggested independently by Dewdney [Dew74] and Beineke and Pippert [BP77].

Let us agree that the vertex set is $[n] = \{1, \ldots, n\}$. Fix the *edge size* k and the *overlap size* m, $0 \le m \le k-1$. We refer to k-subsets and m-subsets of $[n]$ as *edges* and *laps* respectively. A non-empty k-graph without isolated vertices is called a (k,m)*-tree* if we can order its edges, say $E_1, \ldots, E_e$, so that for every i, $2 \le i \le e$, there is i', $1 \le i' < i$, such that $|E_i \cap E_{i'}| = m$ and $(E_i \setminus E_{i'}) \cap \left(\cup_{j=1}^{i-1} E_j\right) = \emptyset$. In other words, we start with a single edge and can consecutively affix a new edge along an m-subset of an existing edge.

Thus, a (k,m)-tree with e edges has $n = e(k-m)+m$ vertices and its edges cover $f = e\left(\binom{k}{m} - 1\right) + 1$ laps. For example, a $(k,0)$-tree consists of disjoint edges.

The problem of counting $(m+1,m)$-trees which are known in the literature as m*-trees*, received great attention and was completely settled by Beineke and Pippert [BP69] and Moon [Moo69]. This extends the celebrated theorem of Cayley [Cay89] as, clearly, 1-trees correspond to usual (Cayley) trees. Later, different bijective proofs for m-trees appeared as well, see [RR70, Foa71, GI75, ES88, Che93].

Here we enumerate (k,m)-trees. In fact, a considerable difficulty was to guess the right formula. When we had a plausible conjecture, we tried to prove it inductively by writing a recurrence relation. We were rather fortunate: the result reduced to the identity proved by Beineke and Pippert [BP69, Lemma 2]. This enabled us to write a short inductive proof, published in [Pik99c], which is presented in Section 19.

Of course, a bijective proof (that is, a correspondence between the set of trees to count and some simple set) is a far more satisfactory answer. (For example, a bijective proof may allow us to generate one by one all trees or to count the number of trees satisfying some given property.) In Subsection 20.2 we exhibit an explicit bijection between the set of rooted vertex labelled trees of given size and a trivially simple set; it is based on the ideas of Foata [Foa71] which are presented in Subsection 20.1. The knowledge of the actual formula was essential, as otherwise we would have had little idea what and how to biject.

In fact, this method (based on Foata's bijection) can be applied to enumerate bijectively other tree-like structures. For example, we can enumerate so called *k-gon trees*, a structure studied in [CL85, Whi88, Pen93, KT96]. In order not to repeat the same portions of proof twice, we present a more general result including both (k, m)-trees and k-gon trees as partial cases.

In Subsection 20.3 we consider the question whether our bijection can count edge labelled trees. We present a construction for 2-graphs only, which in fact answers a question posed by Cameron [Cam95]. This question was motivated by the possibility that such a bijection might simplify some of his enumeration results (or proofs) from [Cam95]. However, although we answered Cameron's question, we were not able to improve [Cam95]. Please refer to Subsection 20.3 for further details.

19 Inductive Proof

Let $T_{km}(e)$ denote the number of distinct (k, m)-trees on $[n]$ with e edges, $n = e(k-m)+m$, and let $R_{km}(e)$ count the trees *rooted* at the lap $[m]$, that is, those trees for which $[m]$ is covered by some edge.

Theorem 75 *Given integers k, m, e with $0 \le m \le k-1$ and $e \ge 1$, let $n = e(k-m)+m$, $l = \binom{k}{m}$ and $f = e(l-1)+1$. Then the number of different (k, m)-trees on $[n]$ equals*

$$T_{km}(e) = \frac{n!f^{e-2}}{e!m!\left((k-m)!\right)^{e}}. \tag{98}$$

Proof. As in Beineke and Pippert [BP69], to prove the theorem, we write down a recurrence relation for $T_{km}(e)$ and then verify that (98) does satisfy the relation. Let us agree that $T_{km}(0) = R_{km}(0) = 1$.

Counting in two different ways the number of pairs (H, L), where H is a (k, m)-tree on $[n]$ rooted at $L \in [n]^{(m)}$, we obtain

$$\binom{n}{m} R_{km}(e) = f \cdot T_{km}(e). \tag{99}$$

Next, consider the following method for constructing trees. Select an edge $E \in [n]^{(k)}$ and label by $L_1, \ldots, L_l$ the laps of E. Represent $e-1$ as a sum of l non-negative integers, $e-1 = e_1 + \ldots + e_l$. Partition $[n] \setminus E$ into sets $X_1, \ldots, X_l$ of sizes $e_1(k-m), \ldots, e_l(k-m)$ respectively. On each $L_i \cup X_i$ build a (k, m)-tree H_i rooted at L_i, $i \in [l]$. Clearly, the union of all H_i's plus the edge E forms a (k, m)-tree and every such tree is obtained exactly e times. Therefore, by (99), we obtain

$$\begin{aligned} eT_{km}(e) &= \binom{n}{k} \sum_{\mathbf{e}} \frac{(n-k)!}{(e_1(k-m))! \ldots (e_l(k-m))!} \prod_{i=1}^{l} R_{km}(e_i) \\ &= \frac{n!}{k!} \sum_{\mathbf{e}} \prod_{i=1}^{l} \frac{m!(e_i(l-1)+1) T_{km}(e_i)}{(e_i(k-m)+m)!}, \end{aligned} \tag{100}$$

where $\sum_{\mathbf{e}}$ denotes the summation over all representations $e - 1 = e_1 + \ldots + e_l$ with non-negative integer summands.

Clearly, formula (98) gives correct values for $e = 0$. Also, the substitution of (98) into the both sides of (100) gives (after routine cancellations)

$$l(e(l-1)+1)^{e-2} = \sum_{\mathbf{e}} \frac{(e-1)!}{e_1! \ldots e_l!} \prod_{i=1}^{l} (e_i(l-1)+1)^{e_i - 1}.$$

The last identity (in slightly different notation) was established by Beineke and Pippert [BP69, Lemma 2], which proves our theorem by induction. ∎

Corollary 76 *The number of vertex labelled m-trees on n vertices, $n > m \geq 1$, is $T_{m+1,m}(n-m) = \binom{n}{m}(mn - m^2 + 1)^{n-m-2}$.* ∎

20 Bijective Proofs

20.1 Foata's Bijection

Given disjoint finite sets A, B, C and a surjection $\gamma : B \to A$, a function $f : A \to B \cup C$ is called *cycle-free* if for every $b \in B$ the sequence $(f \circ \gamma)^i(b)$

eventually terminates at some $c \in C$. Foata [Foa71, Theorem 1] exhibited a bijection between $F(A, B, C, \gamma)$, the set of cycle-free functions, and the set of functions $g : A \to B \cup C$ such that $g(a_1) \in C$, some beforehand fixed $a_1 \in A$; this implies

$$|F(A, B, C, \gamma)| = |C|(|B| + |C|)^{|A|-1}. \tag{101}$$

We briefly describe a simpler construction than that in [Foa71]. Fix some ordering of A. Let $f \in F(A, B, C, \gamma)$. Let $Z = (z_1, \ldots, z_s)$ denote the increasing sequence of the elements in $A \setminus \gamma(f(A))$. (For convenience we assume that $\gamma(c) = c$, $c \in C$.) We build, one by one, s sequences $\delta_1, \ldots, \delta_s$ composed of elements in $B \cup C$. Having constructed sequences $\delta_1, \ldots, \delta_{i-1}$, let $m_i \geq 0$ be the smallest integer such that $(f \circ \gamma)^{m_i}(f(z_i))$ either belongs to C or occurs in at least one of $\delta_1, \ldots, \delta_{i-1}$. We define (mind the order)

$$\delta_i = \left((f \circ \gamma)^{m_i}(f(z_i)),\ (f \circ \gamma)^{m_i-1}(f(z_i)),\ \ldots,\ f(z_i)\right). \tag{102}$$

One can check that Z is non-empty if A is, every m_i exists, and δ, the juxtaposition product of the s sequences $\delta_1, \ldots, \delta_s$, contains $|A|$ elements. (In fact, δ is but a permutation of $(f(a))_{a \in A}$.) The obtained sequence δ of $|A|$ elements of $B \cup C$, which starts with an element of C, corresponds naturally to the required function $g : A \to B \cup C$.

Conversely, given g (or δ), we can reconstruct Z which consists of the elements of $A \setminus \gamma(g(A))$. Then, exactly $s = |Z|$ times, an element of δ either belongs to C or equals some preceding element. These s elements mark the beginnings of $\delta_1, \ldots, \delta_s$. Now we can restore the required f by (102). To establish (101) completely, one has to check easy details.

20.2 H-Built-Trees

Adopting the ideas of Foata [Foa71], we present a bijective proof of (98). Our method can enumerate some other tree-like structures. For example, we can find a bijection for vertex labelled *k-gon trees* (also known as *cacti* or *trees of polygons*), a structure that appears in [CL85, Whi88, Pen93, KT96].

We define a *k-gon tree* inductively. A k-gon (that is, a k-cycle) is a k-gon tree. A k-gon tree with $g+1$ k-gons is obtained from a k-gon tree with g k-gons by adding $k-2$ new vertices and a new k-gon through these vertices and an already existing edge. Thus, a k-gon tree is a (usual) 2-graph; if we have g k-gons, then it has $e = g(k-1)+1$ edges and $n = g(k-2)+2$ vertices.

In order not to repeat the same portions of proof twice, we present the following, more general, result which includes (k,m)-trees and k-gon trees as partial cases.

Let H be any m-graph on $[k]$. An *H-built-tree* $(T, \{H_1, \ldots, H_e\})$ consists of a usual (k,m)-tree T with edges $E_1, \ldots, E_e$ plus H-graphs H_i on E_i, $i \in [e]$, such that if $E_i \cap E_j$ is a lap (that is, has size m), then it is an edge of both H_i and H_j, for any distinct $i, j \in [e]$. Let $n = e(k-m)+m$ be the total number of vertices and let

$$f = \left|\cup_{i\in[e]} E(H_i)\right| = e(l-1)+1,$$

where $l = e(H)$. Also, let $\mathcal{R}_H$ be the set of distinct H-graphs on $[k]$ rooted at $[m]$, that is, containing $[m]$ as an edge. Clearly,

$$|\mathcal{R}_H| = \frac{k!l}{\binom{k}{m}|\mathrm{Aut}(H)|}.$$

An H-built-tree is *rooted* on an m-set L if $L \in \cup_{i\in[e]} E(H_i)$.

Theorem 77 *There is a bijection between the set Y of H-built-trees on $[n]$ rooted at $[m]$ and the set*

$$Z = F(A, B, C, \gamma) \times \prod_{i=1}^{e} (X_i \times \mathcal{R}_H),$$

where $A = [e]$, $B = [e] \times [l-1]$, $C = \{[m]\}$, γ is the coordinate projection $B \to A$, and $X_i = [\binom{(k-m)(e-i+1)-1}{k-m-1}]$. In particular,

$$|Y| = f^{e-1}\left(\frac{k!l}{\binom{k}{m}|\mathrm{Aut}(H)|}\right)^e \prod_{i=1}^{e}\binom{(k-m)(e-i+1)-1}{k-m-1}.$$

Proof. Given an H-built-tree T rooted at $[m]$, order its edges $E_1, \ldots, E_e$ so that $[m] \in E(H_1)$ and each E_i, $i \in [2, e]$, shares a lap with some E_j, $j < i$.

Correspond an edge E_i to the lap $g'(E_i) = E_i \cap \cup_{j=1}^{i-1} E_j$, $2 \le i \le e$. (We agree that $g'(E_1) = [m]$.) Call the set $f(E_i) = E_i \setminus g'(E_i)$ the *free part* of E_i; the free parts partition $[n] \setminus [m]$. Clearly, these definitions of g' and f do not depend on the particular ordering.

Relabel the edges by $D_1, \ldots, D_e$ so that $d_i = \min f(D_i)$ increases; let H_i' denote the corresponding H-graph on D_i. Label, in the colex order, all edges (laps) of H_i' but $g'(D_i) \in E(H_i')$ by (i,j), $j = 1, \ldots, l-1$. Note that now we have indexing of the edges of T by A, namely $(D_i)_{i \in A}$, and of the laps of T by $B \cup C$. Let $g : A \to B \cup C$ be the map corresponding to g'. A moment's thought reveals that g is cycle-free.

Repeat the following for $i = 1, \ldots, e$. Index, in the colex order, the $(k-m-1)$-subsets of $(\cup_{j=i}^{e} f(D_j)) \setminus \{d_i\}$ by the elements of X_i and let $x_i \in X_i$ be the index corresponding to $f(D_i) \setminus \{d_i\}$. Consider the bijection $h : D_i \to [k]$ such that h is monotone on $g'(D_i)$ and $f(D_i)$ which are respectively mapped onto $[m]$ and $[m+1, k]$. Let $r_i \in \mathcal{R}_H$ be the image of H_i' under h.

Now, $(g, x_1, r_1, \ldots, x_e, r_e) \in Z$ is the 'code' of $T \in Y$.

Conversely, given an element $(g, x_1, r_1, \ldots, x_e, r_e) \in Z$ we can consecutively reconstruct the sequence $(d_i, f(D_i))$, $i = 1, \ldots, e$. Indeed, d_i is the smallest element of $V = [n] \setminus ((\cup_{j=1}^{i-1} f(D_j)) \cup [m])$ while $f(D_i) \setminus \{d_i\}$ is the x_ith $(k-m-1)$-subset of $V \setminus \{d_i\}$. For $i \in A$ with $g(i) \in C$, we have $D_i = [m] \cup f(D_i)$ and (knowing $g'(D_i) = [m]$ and $f(D_i)$), we can determine H_i' from r_i; then we can recover the lap corresponding to $(i,j) \in B$ as the jth lap of $E(H_i') \setminus \{[m]\}$, $j \in [l-1]$.

Likewise, we can reconstruct all information about D_i for any $i \in A$ with $g(i)$ being already associated with a lap. As f is cycle-free, all edges are eventually identified, producing $T \in Y$.

A plain verification shows that we have indeed a bijective correspondence between Y and Z. ∎

It is trivial to check that if a union of K_k^m-graphs can be formed into a K_k^m-built-tree, then the latter is uniquely defined. Hence, the number of vertex labelled (k,m)-trees equals the number of K_k^m-built-trees. Now, $|\mathcal{R}_{K_K^m}| = 1$, $|Y| = R_{km}(e)$, and we can easily deduce formula (98).

Similarly, k-gon trees are in bijective correspondence with C_k-built-trees. We have $|\mathcal{R}_{C_k}| = (k-2)!$, so we obtain that there are

$$(g(k-1)+1)^{g-1}((k-2)!)^g \prod_{i=1}^{g} \binom{(k-2)(g-i+1)-1}{k-3}$$

rooted k-gon trees with g k-gons, which implies the following result.

Corollary 78 *The number of vertex labelled k-gon trees with g k-gons is*

$$\frac{(g(k-2)+2)!(g(k-1)+1)^{g-2}}{2(g!)}, \quad k \geq 3. \blacksquare$$

20.3 Edge Labelled Trees

Cameron [Cam95] enumerates certain classes of what is called there *two-graphs*: reduced, 5-free, and $(5,6)$-free and presents their connections to Coxeter groups of graphs. Please refer to his work for all definitions and details. Also, he defines, for a given (Cayley) tree T, the equivalence relation $\cong$ on its edges which is the smallest one such that two edges are related if they intersect at a vertex of degree 2 in T. For example, T is series-reduced (that is, T does not contain a vertex of degree 2) if and only if $\cong$ is the identity relation.

Cameron had to count the number S_n of trees with n edges with labelled $\cong$-classes. He found the following formula ([Cam95, Proposition 3.5(a)]):

$$S_n = \sum_{k=1}^{n} S(n,k) \frac{1}{k+1} \sum_{j=0}^{k-1} (-1)^j \binom{k+1}{j} \binom{k-1}{j} j!(k-j+1)^{k-j-1}, \quad (103)$$

where $S(n,k)$ is the Stirling number of the second kind: the number of partition of an n-set into k non-empty parts. The sequence (S_n) starts as $1, 1, 2, 8, 52, \ldots$ and probably cannot be represented in a simple form but, of course, one can try to simplify (103).

Cameron [Cam95] asks the following question.

Problem 79 (Cameron) *Describe a constructive bijection between edge labelled trees and edge Prüfer codes, not going via vertex labellings. Describe the equivalence relation $\cong$ in terms of this code.*

The motivation for the problem was apparently that such a code might simplify (103). Although we answer here this question, we have not so far been able to simplify (103) or its proof from [Cam95]. But anyway, let us describe our construction. Of course, we use Foata's [Foa71] bijection for cycle-free functions.

Let $e_1, \ldots, e_n$ be the edges. Suppose $e_1 = \{a, b\}$; this edge will play a special role. Let $A = B = \{e_2, \ldots, e_n\}$, $C = \{a, b\}$ and $\gamma : B \to A$ be the identity function. Let us correspond an $f \in F(A, B, C, \gamma)$ to a given tree T. Each edge e can be connected to e_1 by the unique path in T. If e is incident to e_1, then let $f(e)$ be equal to their common vertex; otherwise, let $f(e)$ be the first edge on the path from e to e_1. This gives a correspondence between twice the number of edge-labelled trees (we can label the two vertices of e_1 by a and b in two different ways) and $F(A,B,C,\gamma)$. Foata's bijection shows that $|F(A,B,C,\gamma)| = |C|(|A|+|C|)^{|A|-1}$, which implies, as desired, that the number of edge-labelled trees with n edges is $(n+1)^{n-2}$.

Of course, the code is rather simple; we describe briefly only one direction. A code δ is a sequence of length $n-1$ consisting of elements in $\{a,b,e_2,\dots,e_n\}$ and staring with a or b. The set $Z \subset \{e_2,\dots,e_n\}$ of edges which do not occur in the sequence consists of leaves. (If a or b does not occur, then e_1 is also a leaf.) Clearly, an element of δ equals either a or b or some previously occurring element exactly $z = |Z|$ times. Cut δ before each such element; we have z pieces $\delta_1,\dots,\delta_z$. Append the ith element z_i of Z to the end of δ_i to obtain δ_i', $i \in [z]$.

The reversed sequence δ_i' describes the initial segment P_i' of the path P_i from the element $z_i \in Z$ to e_1 until it hits e_1 or some previous path P_j, $i \in [z]$. Clearly, this determines some tree.

This bijection corresponds to every edge-labelled tree two codes, one starting with a and the other—with b. To make this correspondence one-to-one, we consider only a half of the codes, e.g. those starting with a.

How can we read the $\cong$-relation from δ? First, let $\cong'$ be the minimal equivalence relation on $\{a,b,e_2,\dots,e_n\}$ such that $e_i \cong' e_j$ if e_i and e_j intersect at a vertex of degree 2, $2 \le i < j \le n$, and $x \cong' e_i$ if x is a degree-2 vertex incident to e_i, $x \in \{a,b\}$, $i \in [2,n]$. (Informally, we cut e_1 in its middle and take the usual $\cong$-relation on the both created components separately.) Clearly, $\cong$ is obtained from $\cong'$ by identifying a and b into a single element e_1, so let us indicate how to determine the latter relation.

Take any maximal contiguous subsequence $S \subset \delta$ consisting of elements that occur in δ exactly once. Clearly, S lies entirely within some δ_i and $S \cup \{y\}$ is a $\cong'$-equivalence class, where y is the symbol following S in δ_i'. Conversely, it is easy to check that all non-trivial $\cong'$-classes are obtained this way, as required.

This answers Problem 79. Unfortunately, I do not see how this description can simplify Cameron's formula (103).

Remark. We do not know any bijection enumerating edge labelled (k,m)-trees for $k \ge 3$.

Part V

Large Degrees in Subgraphs

21 Introduction

All research carried in this part revolves around the following conjecture of Erdos [Erd81] which is disproved here.

Erdős [Erd81], see also e.g. [Chu97, Erd99], conjectured that for $n \geq 3$ any graph with $\binom{2n+1}{2} - \binom{n}{2} - 1$ edges is a union of a bipartite graph and a graph with maximum degree less than n. This value arises from the consideration of $P_{n+1,n}$ which does not admit the above representation. ($P_{m,n} = K_m + E_n$ has $m+n$ vertices of which m vertices are connected to every other vertex.)

In the arrowing notation the latter statement reads "$P_{n+1,n} \to (K_{1,n}, \mathcal{C}_{\mathrm{odd}})$": for any blue-red colouring of the edge-set of $P_{n+1,n}$ we necessarily have either a blue star $K_{1,n}$ or a red cycle of odd length. (By $\mathcal{C}_{\mathrm{odd}}$ we denote the family of odd cycles.) Thus the conjecture states that $\hat{r}(K_{1,n}, \mathcal{C}_{\mathrm{odd}}) = e(P_{n+1,n})$ and, if true, would give the same value for the size Ramsey number $\hat{r}(K_{1,n}, K_3)$, since certainly $\hat{r}(K_{1,n}, K_3) \geq \hat{r}(K_{1,n}, \mathcal{C}_{\mathrm{odd}})$ and in fact $P_{n+1,n} \to (K_{1,n}, K_3)$.

We show, however, that both these size Ramsey numbers grow as n^2 plus a term of order $n^{3/2}$. (Actually, the conjecture fails for all $n \geq 5$.) More precisely, our main result is the following.

Theorem 80

$$\hat{r}(K_{1,n}, K_3) < n^2 + \sqrt{2}\, n^{3/2} + n, \qquad \textit{for } n \geq 1, \tag{104}$$

$$\hat{r}(K_{1,n}, \mathcal{C}_{\mathrm{odd}}) > n^2 + 0.577\, n^{3/2}, \qquad \textit{for sufficiently large } n. \tag{105}$$

In [FRS97, Section 1] it is asked whether the conjecture is true for graphs with (at most) m vertices. Faudree (for a proof see [ERSS96]) showed this is the case for $m = 2n+1$. Our construction can beat $P_{n+1,n}$ on $3n+1$ vertices. Perhaps $P_{n+1,n}$ is extremal for graphs with $2n$ plus few more vertices, but even for $2n+2$ vertices we do not know whether this is true.

Some previous attempts to prove Erdős' conjecture resulted in new interesting directions of research; here we investigate also some of these questions.

Erdős and Faudree [EF99] consider the related problem of determination of the minimal size of a graph G such that if G is a union of two graphs, one having maximal degree less than n, then the other contains all odd cycles C_m with $3 \le m \le n-3$. Here we demonstrate a graph G of size $(1+\varepsilon)n^2$, for any given constant $\varepsilon > 0$, such that, for any blue-red colouring of G without a blue $K_{1,n}$, we have red cycles of all lengths (odd and even) between 3 and cn, where $c = c(\varepsilon) > 0$ does not depend on n.

For positive integers n, k, j with $k \ge j$, Erdős, Reid, Schelp and Staton [ERSS96] consider the property $\mathcal{M}(n,k,j)$ which is defined as follows. A graph G belongs to $\mathcal{M}(n,k,j)$ if it has $n+k$ vertices and for every $(n+j)$-set $A \subset V(G)$ we have $\Delta(G[A]) \ge n$. (That is, the maximal degree of the subgraph of G spanned by A is at least n.) The problem is to compute

$$m(n,k,j) = \min\{e(G) : G \in \mathcal{M}(n,k,j)\}.$$

Erdős et al [ERSS96, Conjecture 1] conjectured that for any $n \ge k \ge j \ge 1$ and $n \ge 3$, we have

$$m(n,k,j) = (k-j+1)n + \binom{k-j+1}{2}. \tag{106}$$

This value arises from the consideration of $P_{k-j+1,n} \sqcup E_{j-1}$. Erdős et al [ERSS96, Theorem 3] proved that (106) is true if $j = 1$ or if $j \ge 2$ and

$$n \ge \max\left(j(k-j),\ \tbinom{k-j+2}{2}\right). \tag{107}$$

In Section 23 we demonstrate a constructive counterexample to (106) if $n \le (j-2)(k-j)$. On the other hand, we show that the formula (106) is true if

$$n \ge \max\left(\left(j+\tfrac{1}{2}\right)(k-j) + \tfrac{j+k}{4j-2}, 14\right),$$

which is an improvement on (107) for $j \lessapprox k/3$. This shows that the value $j(k-j)$ is roughly the threshold on n when the obvious construction suggesting (106) fails to be extremal. Some other constructions are presented.

Another function whose study was motivated by Erdős' conjecture is as follows. Let $\mathcal{B}(n,m)$ consist of all graphs such that, for any partition $V(G) = A \cup B$, either $\Delta(G[A]) \ge n$ or $\Delta(G[B]) \ge m$ (or both). We are interested in the *bisplit function* $b(n,m) = \min\{e(G) : G \in \mathcal{B}(n,m)\}$. Clearly, $b(n,n) =$

$\hat{r}(K_{1,n}, \mathcal{C}_{\mathrm{odd}})$ is precisely the function investigated in Erdős' conjecture, which was the original motivation for introducing the 'off-diagonal' numbers $b(n, m)$.

In Subsection 24.1 we present a simple argument giving a lower bound on $b(n, m)$, any n, m, and a construction of $G \in \mathcal{B}(n, m)$ (which obviously gives an upper bound) which together compute the function asymptotically when $m = \min(n, m)$ is large:

$$b(n, m) = 2nm - m^2 + o(m)n. \tag{108}$$

Concerning small values of m, not much is known. Of course, the bounds of Subsection 24.1 are applicable here, but the error term is not negligible if m is bounded. Namely, we obtain that, for any fixed $m \geq 1$, the numbers $b(n, m)$, $n \in \mathbb{N}$, lie between two functions linear in n with slopes $2m + 1$ and $2m + \sqrt{2m} + \frac{5}{2}$.

We prove that $b(n, 1) = 4n - 2$ for $n \geq 8$ (and characterize all extremal graphs) and that $b(n, 2) = 6n + O(1)$. As the reader will see the proofs are rather lengthy and require consideration of many cases. This indicates that the computation of $\lim_{n\to\infty} b(n, m)/n$ for any fixed m (if the limit exists) is perhaps a hard task.

22 Triangle-vs-Star Size Ramsey Number

Here we will prove the bounds on $\hat{r}(K_{1,n}, K_3)$ stated in the introduction.

22.1 Upper Bound

Proof of (104). We provide an explicit construction of a $(K_{1,n}, K_3)$-arrowing graph G.

Take any representation $n = k_1 + \ldots + k_m$ and let G be the disjoint union of $P_{k_i,n}$, $i \in [m]$, plus a vertex x connected to everything else. Consider any blue-red colouring of $E(G)$ without a blue $K_{1,n}$. Among $n(m+1)$ edges incident to x there are at least $mn+1$ red ones. By the pigeon-hole principle, x sends at least $n + 1$ red edges to some $P_{k_j,n}$, say $\{x, y_i\}$, $i \in [0, n]$, of which at least one must be incident to a vertex of $K_{k_j} \subset P_{k_j,n}$, say y_0. But of n edges $\{y_0, y_i\} \in E(G)$, $i \in [n]$, one is necessarily red and creates a red triangle whose third vertex is x. Hence, $G \to (K_{1,n}, K_3)$.

We have $e(G) = (m+n+1)n + \sum_{i\in[m]} \binom{k_i}{2}$. To minimize $e(G)$ we take the k_i's nearly equal; so they are essentially uniquely determined by m. Any value of m we choose will give some upper bound for $\hat{r}(K_{1,n}, K_3)$. Choose m so that $n = 2m^2 + r$, where $|r| \leq 2m$. So, for example, when $n = 2m^2 - 2m$ we could choose either m or $m-1$. We believe, though we do not prove, that such a choice of m is optimal. The verification of (104) is now best split into four cases. For example, for $0 \leq r \leq m$ we have $m - r$ times $k_i = 2m$ and r times $k_i = 2m+1$. Routine simplifications show that

$$2n^3 - (e(G) - n^2 - n/2 - r/2)^2 = 3m^2r^2 + 2r^3 \geq 0,$$

which implies (104). The other cases can be verified similarly. ∎

One can check that the bound (104) gives strictly better values than $\binom{2n+1}{2} - \binom{n}{2}$ for all $n \geq 6$. In fact, Erdős' conjecture fails also for $n = 5$ when the representation $n = 2 + 3$ produces a graph with 44 edges.

We do not know any example beating our construction, which therefore might be an extremal one, but we do not dare to make any conjecture yet. It is surprising that a counterexample was not found earlier. An explanation might be that $P_{n+1,n}$ is perhaps extremal among all $(K_{1,n}, \mathcal{C}_{\mathrm{odd}})$-arrowing graphs with few vertices; as shown by Faudree (for a proof see [ERSS96]) this is the case for graphs of order $2n+1$. Note that we can beat $P_{n+1,n}$ using $3n+1$ vertices for $n \geq 5$: take $m = 2$ in our construction.

22.2 Lower Bound

In this section we suppose on the contrary to (105) that there is a $(K_{1,n}, \mathcal{C}_{\mathrm{odd}})$-arrowing graph G with at most $n^2 + 0.577\, n^{3/2}$ edges and try to derive a contradiction for large n.

Instead of 2-colourings of $E(G)$ we find it more convenient to operate with 2-partitions of $V(G)$. Thus our assumption on G states that

$$\max\{\Delta(G[A]), \Delta(G[B])\} \geq n$$

for any partition $V(G) = A \cup B$.

The following simple argument, which we call the *greedy algorithm*, shows that any $A \subset V(G)$ spans at least $n(n - |\overline{A}| + 1)$ edges, where $\overline{A} = V(G) \setminus A$.

Indeed, inductively let x_i be any vertex (if exists) of degree at least n in $G[A \setminus \{x_1, \ldots, x_{i-1}\}]$. Let $X = \{x_1, \ldots, x_k\} \subset A$ be the set eventually obtained. By definition, $\Delta(G[A \setminus X]) < n$. But then $\overline{A} \cup X$ contains at least $n+1$ vertices (to allow a vertex of degree n), that is, $k \geq n - |\overline{A}| + 1$, and the claim follows.

Taking $A = V(G)$ we obtain $e(G) \geq n^2 + n$. We will add an $n^{3/2}$-term to this trivial bound by using a probabilistic argument. But before we can apply it, we have to fiddle a lot with the greedy algorithm in order to gain some structural information about G.

Let us introduce some notation first. By $d_A(x) = |A \cap \Gamma(x)|$ we denote the number of neighbours of x lying in A, $x \in V(G)$, $A \subset V(G)$. Also let $L = \{x \in V(G) : d(x) \geq n\}$, $l = |L| - n$ and $e(G) = n^2 + c_g n^{3/2}$. Thus we assume that $c_g \leq 0.577$ and in fact, by adding edges to G, that $c_g = 0.577 + o(1)$.

Lemma 81 $l \leq c_g n^{1/2} + O(1)$.

Proof. Apply a modified greedy algorithm. Set initially $A = C = \emptyset$ and $B = V(G)$. These three sets will always partition $V(G)$.

Repeat the following as long as possible or until $|A| = n+1$. Take a vertex $x \in B$ (if exists) with $d_B(x) \geq n$ and move it to A; colour aqua all edges connecting x to B. Then for every such x do the *n-check*, that is, move to C all vertices in $B \cap L$ whose $B \cup C$-degree is now smaller than n, that is, equals $n-1$. (Thus before we proceed with another x we ensure that a vertex $z \in L \setminus A$ belongs to B if and only if $d_{B \cup C}(z) \geq n$.)

When we stop we have $a + c \geq n+1$, where a, b, c are the cardinalities of the eventual sets A, B, C. Indeed, if $a < n+1$ then $\Delta(G[B]) < n$ so $\Delta(G[A \cup C]) \geq n$ and the claim follows.

The number of aqua edges is $e_a \geq an$. Call non-aqua edges incident to C cyan. Every vertex in C is incident to exactly $n-1$ cyan edges; hence we have $e_c \geq c(n-1) - \binom{c}{2}$ cyan edges.

By applying our usual greedy algorithm to $B \cup C$ we obtain that there is a set $Y = \{y_1, \ldots, y_{n+1-a}\} \subset B \cup C$ such that each y_i has at least n neighbours in the complement of $A \cup \{y_1, \ldots, y_{i-1}\}$. Clearly, Y must be disjoint from C, that is, $Y \subset B$. We have $e_y \geq (n+1-a)n$ edges between Y and $C \cup B$; colour all these edges yellow. (Some edges may be yellow and cyan simultaneously.) Finally, each vertex in $R = L \cap (B \setminus Y)$ has degree in $B \cup C$ at least n (otherwise it would have been moved to C earlier). Hence R is incident to $e_r \geq r(n-|Y|-c)-\binom{r}{2}$ edges lying within $B\setminus Y$, where $r=|R|$; call them red edges.

We claim that $c=o(n)$. Suppose not. As $e_a+e_y>n^2$, the number of cyan-only edges is $o(n^2)$ and the average yellow-degree of $x\in C$ is $n+o(n)$; hence $|Y|\geq n+o(n)$. Now $|C|\geq |Y|$ because $a+c\geq n+1=a+|Y|$, so $|C|\geq n+o(n)$. But $C\cup Y\subset L$ and $|L|\leq 2n+o(n)$ by the handshaking lemma. Therefore $c=n+o(n)$, $a=o(n)$, $r=o(n)$ and all but $o(n^2)$ edges lie between C and Y. But consider partition $V(G)=V_1\cup V_2$ obtained by placing in V_1 all of $A\cup R$, $n/3$ vertices from C, $n/3$ vertices from Y and all ($=o(n)$) vertices from C (and resp. from Y) which have in G at least $n/6$ neighbours outside Y (resp. outside C). As $|V_1|=2n/3+o(n)$ some $x\in V_2$ satisfies $d_{V_2}(x)\geq n$. But x necessarily belongs to $Y\cup C$, say $x\in C$, and can have at most $|Y\cap V_2|+n/6\leq 5n/6+o(n)$ V_2-neighbours, which is a contradiction proving $c=o(n)$.

Using the above lower bounds on e_a, e_c, e_y and the inequality $a\geq n-c+1$ we obtain

$$\begin{aligned} e(G) &= n^2+c_g n^{3/2} \geq e_a+e_c+e_y-(n-a+1)c \\ &\geq n^2+n-\tfrac{c^2+3c}{2}+ac \geq n^2+n+\tfrac{-3c^2+c(2n-1)}{2}. \end{aligned}$$

Solving this (quadratic in c) inequality we obtain that necessarily $c<c_g n^{1/2}$ for large n as c cannot be bigger than the larger root $2n/3+o(n)$.

Writing $e(G)\geq e_a+e_c+e_y-(n-a+1)c+e_r$ and substituting $a\geq n-c+1$ everywhere (as the total coefficient of a is positive) we obtain

$$c_g n^{3/2} \geq \tfrac{-r^2+r(2n+1)}{2}+cn-2cr+O(n).$$

The larger root of this quadratic in r inequality is $2n+o(n)$, but $r\leq n+o(n)$ since $a=n+o(n)$ and $a+r\leq |L|$. So we conclude that $l-1=c+r\leq c_g n^{1/2}+O(1)$ as required. ∎

Now let us try to derive a final contradiction.

Proof of (105)**.** Let $x_{\max}$ be a vertex of maximal degree $\Delta(G)=c_m n^{3/2}$. The greedy algorithm shows that $e(G)\geq n^2+\Delta(G)$, that is, $c_m\leq c_g$. Let $c'=((4+c_g^2)^{1/2}-c_g)/2$ and $c_f=1.732$.

We apply a version of the greedy algorithm. Set initially $A=C=\emptyset$ and $B=V(G)$.

At Stage 1 move to A, one by one and as long as possible, a vertex $x \in B$ with $d_{B\setminus L}(x) \geq n-l$ and $d_{B\cup C}(x) \geq n$. After x was moved do the n-check, that is, move to C all vertices $y \in B \cap L$ with $d_{B\cup C}(y) < n$. We may assume that we were selecting $x \in B$ so that $d_G(x)$ was non-increasing. Let A_1 be the set of vertices moved to A at Stage 1, $F = \{x \in A_1 : d_G(x) \geq n + c_f n^{1/2}\}$ and $a_f = |F|/n$. By Lemma 81 we have $l \leq c_g n^{1/2} + o(1)$, so the number of edges incident to F is at least

$$\sum_{x\in F} d(x) - \sum_{x\in F} \frac{d(x)-n+l}{2} \geq a_f n^2 + a_f \frac{c_f - c_g}{2} n^{3/2} + o(n^{3/2}).$$

At Stage 2 move to A, one by one and as long as possible, any vertex $x \in B$ having at least $n + c'n^{1/2}$ neighbours in $B \cup C$ and for every such x do the n-check as in Stage 1.

At Stage 3 we repeat the following until $B \cap L = \emptyset$. Take $x \in B \cap L$. As long as $d_{B\cup C}(x) \geq n$ move to A some x-neighbour $y \in B \cap L$ (note that $d_{B\cup C}(y) \geq n$) and perform the n-check. Such y necessarily exists as x has fewer than $n-l$ neighbours in $B \setminus L$ while $|C| \leq l$. (The latter inequality is true because if $|C| > l$ at some moment then continuing with the standard greedy algorithm applied to $B \cup C$ we find at least $n - |A| + 1$ vertices in $B \cap L$ which contradicts $|L| = n + l$.) Of course, the last n-check moves x itself to C.

Let $a_i n$ (resp. $c_i n^{1/2}$) be the number of vertices moved to A (resp. to C) at the ith Stage. As eventually $\Delta(G[B\cup C]) < n$ we conclude that $a_1 + a_2 + a_3 > 1$. Also $a_3 \leq c'c_3$ as for every x moved to C at Stage 3 we moved at most $c'n^{1/2}$ vertices to A.

Note that the first vertex moved at Stage 1 may be assumed to have degree $\Delta(G) = c_m n^{3/2}$ unless $\Delta(G) = O(n)$. So our algorithm produces the following lower bound on the size of G:

$$e(G) \geq n^2 + \left(c_m + a_f \tfrac{c_f - c_g}{2} + a_2 c' + c_3(1-a_3) + o(1)\right) n^{3/2}. \tag{109}$$

Now using the inequalities $a_3 \leq c'c_3$ (twice) and $0 \leq c_3 \leq c_g + o(1)$ (by Lemma 81 we have $c_3 \leq |C|n^{-1/2} \leq c_g + o(1)$) we obtain from (109) that

$$\begin{aligned} a_2 + a_3 &\leq \frac{c_g - a_f(c_f - c_g)/2 - c_m - c_3 + c'c_3^2}{c'} + c'c_3 + o(1) \\ &\leq \frac{c_g - a_f(c_f - c_g)/2 - c_m}{c'} + \max\left(0, c_g^2 + c'c_g - c_g/c'\right) + o(1). \end{aligned}$$

But our c' satisfies $c_g^2 + c'c_g = c_g/c'$ so the second term disappears.

Choose a set $Y \subset \overline{L}$ by placing each vertex of $\overline{L}$ into Y independently with probability $p = (c_f + 2\varepsilon)n^{-1/2}$, where $\varepsilon > 0$ denotes a small constant.

The number of Y-neighbours of any $x \in L$ has a binomial distribution with expectation at most $pc_m n^{3/2} = (c_f + 2\varepsilon)c_m n$. Hence the probability that say $d_Y(x) > (c_f c_m + 3\varepsilon)n$ is exponentially small in n by Chernoff's bounds [Che52].

Similarly, the expected value of $d_Y(x)$ for $x \in A_1$ is at least $p(n - l) \approx (c_f + 2\varepsilon)n^{1/2}$ and $d_Y(x) < (c_f + \varepsilon)n^{1/2}$ with probability at most $\exp(-cn^{1/2})$ for some constant $c > 0$.

Hence, there exists Y (in fact, almost every choice would do) such that $d_Y(x) \le (c_f c_m + 3\varepsilon)n$ for every $x \in L$ and $d_Y(y) \ge (c_f + \varepsilon)n^{1/2}$ for every $y \in A_1$.

Now consider the partition $V(G) = V_1 \cup V_2$, where $V_1 = (\overline{L} \setminus Y) \cup (A_1 \setminus F)$. Any $x \in A_1 \setminus F$ has at least $(c_f + \varepsilon)n^{1/2} > d(x) - n$ neighbours in Y, so $d_{V_1}(x) < n$. But then $d_{V_2}(x) \ge n$ for some $x \in L \cap V_2$. Hence,

$$n \le |V_2 \setminus Y| + d_Y(x) \le n + l - |A_1| + |F| + (c_f c_m + 3\varepsilon)n,$$

or equivalently

$$a_2 + a_3 + a_f + c_f c_m \ge 1 + \text{error term}, \tag{110}$$

where the error term can be made arbitrarily small by choosing the constant ε small.

Chopping off some terms in (109) we obtain that a_f lies between 0 and $2(c_g - c_m)/(c_f - c_g) + o(1)$. Hence

$$\begin{aligned} a_2 + a_3 + a_f &\le \frac{c_g - a_f(c_f - c_g)/2 - c_m}{c'} + a_f + o(1) \\ &\le \max\left(\tfrac{c_g - c_m}{c'}, 2\tfrac{c_g - c_m}{c_f - c_g}\right) + o(1). \end{aligned} \tag{111}$$

Using the values of c_g and c_f we obtain from (110) and (111) that necessarily

$$\max(0.767 + 0.403\, c_m,\ 0.9992 + 0.0004\, c_m) \ge 1 + o(1),$$

which cannot be satisfied for $0 \le c_m \le 0.577$. ∎

Remark. The constant 0.577 can be improved, even with the present proof. For example, the optimal choice

$$c_f = \min\left(\sqrt{4 + c_g^2},\ c_g + \sqrt{2(c_g - c_m)/c_m}\right),$$

should give (with extra algebraic work) $c_g \geq 0.591$.

Also, after Stage 2 we could apply the algorithm of Lemma 81: we have identified at least $(c_m + a_2(c' - c_g) + a_f\frac{c_f - c_g}{2})n^{3/2}$ 'useless' (from the point of view of Lemma 81) edges, which should bring down the bound on l there. We do not know how much gain this would have given (the calculations get rather messy) but we believe that we have reached a good compromise in the sense that the proof is not too long and the bound is not too bad.

22.3 Cycles of Consecutive Lengths

As we already mentioned, Erdős and Faudree [EF99] study minimum graphs G such that if G is a union of two graphs, one having maximal degree less than n, then the other contains all odd cycles C_m with $3 \leq m \leq n-3$. Here we show, that if we require cycles lengths from 3 to $\Theta(n)$, then we can present a construction with only $(1+\varepsilon)n^2$ edges for any fixed $\varepsilon > 0$.

In the proof below we introduce constants c_1, c_2, and so on. It should not be hard to check that we can always choose c_i (depending on $c_1, \ldots, c_{i-1}$) satisfying all conditions set in the proof. We do not try to optimize the constants.

Theorem 82 *For any fixed $\varepsilon > 0$, there is a graph G with at most $(1+\varepsilon)n^2$ edges such that if $E(G)$ is coloured blue-red without a blue $K_{1,n}$, then we have red cycles of all lengths (even and odd) between 3 and cn for some $c = c(\varepsilon) > 0$ which does not depend on n.*

Proof. Choose integers

$$\begin{aligned} m &= \sqrt{n/2} + O(1), \\ k &= (\sqrt{2} + c_1)\sqrt{n} + O(1), \\ l &= n + c_1 n + O(1), \\ h &= c_1\sqrt{n} + O(1). \end{aligned}$$

Choose k-sets $K_1, \ldots, K_m$, l-sets $L_1, \ldots, L_m$, and an h-set H (all disjoint). Let G consist of all edges intersecting H and of all edges intersecting K_i and lying within $K_i \cup L_i$, $i \in [m]$, that is, $G = K_h + mP_{k,l}$. If $c_1 > 0$ is small, then G has at most $(1+\varepsilon)n^2$ edges.

Consider any blue-red colouring of $E(G)$ without a blue $K_{1,n}$. Let $G' \subset G$ be the red subgraph, let $d'(x)$ be the red degree of $x \in V(G)$, and so on.

Define the bipartite graph F with classes H and $[m]$ as follows; $x \in H$ is connected to $i \in [m]$ if and only if x sends at least $l + c_2\sqrt{n}$ red edges to $K_i \cup L_i$, $c_2 = c_1/2$. Now, the inequality

$$(m - d_F(x))c_2\sqrt{n} + d_F(x)k \geq mk - n + 1,$$

implies that each $x \in H$ has $d_F(x) \geq c_3\sqrt{n}$ neighbours in F.

First, let us show how to find red cycles of all lengths up to $c_4\sqrt{n}$. Choose, any $\{x, i\} \in E(F)$.

In G', we have $d'_{K_i}(x) \geq c_2\sqrt{n}$ and each vertex in $\Gamma'_{K_i}(x)$ has at least $c_1 n + o(n)$ neighbours in $\Gamma'_{K_i \cup L_i}(x)$. (Because the latter set has size $n + c_1 n + O(\sqrt{n})$ while we do not have a blue $K_{1,n}$ in G.) Thus we have $\Theta(n^{3/2})$ red edges within $\Gamma'_{K_i \cup L_i}(x)$. By the theorem of Erdős and Gallai [EG59], we have a red path of length $c_4\sqrt{n}$ there, which together with x creates red cycles of all lengths up to $c_4\sqrt{n}$.

Next, the graph F (which, in fact, has positive density) has a cycle of length $2t = \Theta(\sqrt{n})$ with $4t < c_4\sqrt{n}$ for large n; let it go through vertices $x_1, i_1, , \ldots, x_t, i_t, x_{t+1} = x_1$, where $x_j \in H$ for $j \in [t]$.

To prove the theorem it is clearly enough to show that, for any $j \in [t]$, we can find a red path connecting x_j and x_{j+1} through $K_{i_j} \cup L_{i_j}$ of any length between 2 and $c_5\sqrt{n}$, for some constant c_5.

Consider $X = (\Gamma'(x_j) \cup \Gamma'(x_{j+1})) \cap (K_{i_j} \cup L_{i_j})$. Now, $X \cap K_{i_j}$ has at least $c_2\sqrt{n}$ elements, each being incident to at least $c_1 n + O(\sqrt{n})$ red edges. It is not hard to see that we can find a red cycle C within X of length at least $c_5\sqrt{n}$ intersecting $\Gamma'(x_j) \cap \Gamma'(x_{j+1})$. (The latter set has size $n + c_1 n + O(\sqrt{n})$ and it is incident to almost all red edges lying within X.) It is easy to see that we can additionally require that C has a red cord E. Now, by a simple lemma (which is implicit in Bondy and Simonovits [BS74] and explicit in Verstraëte [Ver99]), $C + E$ contain paths connecting $\Gamma'(x_j) \cap V(C)$ to $\Gamma'(x_{j+1}) \cap V(C)$ (two intersecting sets that cover $V(C)$) of all lengths from 0 to $v(C) - 1$, as required. ∎

Remark. For each $j \in [t]$, we can find a red cycle of any prescribed length between 3 and $c_4\sqrt{n}$ lying within $K_{i_j} \cup L_{i_j} \cup \{x_j\}$. Hence, we can find $t = \Theta(\sqrt{n})$

such vertex disjoint cycles in G. Of course, one can try to prove many other similar results about our graph G. For example, what is $c = c(\varepsilon)$ if ε tends to zero with a given rate as $n \to \infty$, say $\varepsilon = \Theta(1/\sqrt{n})$? But we do not want to build a whole theory out of it: our purpose was to demonstrate that if we allow $(1+\varepsilon)\hat{r}(K_{1,n}, \mathcal{C}_{\mathrm{odd}})$ edges then we can witness much stronger properties.

23 Removing Vertices

In this section we denote $l = k - j \geq 0$. Thus a graph of order $n + k$ belongs to $\mathcal{M}(n, k, j)$ if after the removal of any l vertices the maximal degree is at least n.

23.1 Some Constructions

Here is our counterexample to the conjecture of Erdős, Reid, Schelp and Staton [ERSS96, Conjecture 1].

Example 83 *The formula* (106) *is not true if* $n \leq (j-2)l$.

Proof. Write $n = lq + r$ with $0 \leq r < l$. Let $A = [l+1]$, $y = l + 2$, and $R = [l+3, l+r+2]$, that is, $R \subset X \setminus (A \cup \{y\})$ is a set of size r. Our assumption on n implies that $j \geq q + 2$, that is,

$$n + k - l - r - 2 \geq (l+1)q.$$

Therefore, in $X \setminus (A \cup R \cup \{y\})$ we can choose disjoint q-sets $Q_1, \ldots, Q_{l+1}$. Let our graph G consist of the following edges: $\{f, h\} \in A^{(2)}$ with $|f - h| > 1$ (that is, A spans the complete graph but for a Hamiltonian path), all edges between A and R, all edges connecting $f \in A$ to Q_h with $h \neq f$ and edges $\{f, y\}$, $f \in [2, l]$. Thus all vertices in A have degree $n + l - 1$. It is easy to check that the size of G is by one smaller than the bound given by (106).

We claim that $G \in \mathcal{M}(n, k, j)$. Suppose on the contrary to our claim that we can remove some set L of size l so that the remaining graph has maximal degree less than n. Let x be any vertex in $A \setminus L$ which is not empty as $|A| > l$. As the degree of x should be less than n now, we conclude that x is connected to each vertex in L. Therefore, any vertex in A non-incident to some $x \in A \setminus L$ lies itself in $A \setminus L$. As $\overline{G}[A]$ is connected (it is a path), we conclude that $A \cap L = \emptyset$. But the set of vertices connected (in G) to everything in A is precisely R and it has $r < l$ elements, which is a contradiction. ∎

Remark. In Example 83 we can win a few more edges if n is yet smaller. As before, we let $A = [l+1]$ and $y = l+2$. Suppose that for some p we can squeeze into $[l+3, n+k]$ an r-set R and q-element sets $Q_1, \ldots, Q_{p+1}$ with $r < p$ and $pq + r = n$. To define G, let $\overline{G}[A]$ consist of $h = \lfloor \frac{l+1}{p+1} \rfloor$ vertex-disjoint paths of length $p+1$ each; for every such path $(x_1, \ldots, x_{p+1})$ we connect x_i, $i \in [p+1]$, to everything in R and in Q_j, $j \neq i$, and we connect x_i, $i \in [2, p]$, to y. Also, we add some extra edges so that any vertex of A not on a path has degree $n + l$.

Suppose that $G \notin \mathcal{M}(n, k, j)$, that is, there is an l-set L with $\Delta(G - L) < n$. Let $x_0 \in A \setminus L$. It must lie on a path P. (Otherwise $d_G(x) = n + l$.) Like in Example 83 we argue that P does not intersect L and, in fact, every $x \in P$ is connected to all vertices in L. But the number of vertices connected to the whole of P is $|(L \setminus P) \cup R| < l$, which is a contradiction.

It is easy to see that we have h edges less than in (106), so it is advantageous to choose p as small as possible. The condition we have to satisfy is

$$n + k - l - 2 \geq r + (p+1)q$$

or, equivalently, $j - 2 \geq q = \lfloor n/p \rfloor$. Therefore, we choose $p = \lceil \frac{n}{j-2} \rceil$. Note that our gain compared to (106) is $h = \lfloor \frac{l+1}{p+1} \rfloor$.

Erdős et al [ERSS96] observe that (106) is not true 'when k is very large compared to n.' Here is an example, for any given l and n, giving only a fraction of (106) with k moderately small (starting with $k \geq n + l + 1$).

Example 84 *Suppose that $n + k \geq pn + l + 1$, where p is an integer greater than 1. Take a representation $l + 1 = l_1 + \ldots + l_p$ and let $G = \sqcup_{i \in [p]} (K_{l_i} + E_n)$. Then $G \in \mathcal{M}(n, k, k - l)$.*

Proof. Let $L \subset V(G)$ be any l-set. There must exist $i \in [p]$ such that L intersects the corresponding component C_i in less then l_i vertices. Hence, at least one vertex in K_{l_i} survives and it has at least n neighbours outside L. ∎

23.2 Improving Condition (107)

We can prove the following (which is an improvement of (107) if $j \lessapprox k/3$).

Theorem 85 *Let $j \geq 2$ and $n \geq 14$. Then* (106) *is true if*

$$n \geq \left(j + \frac{1}{2}\right) l + \frac{2j+l}{4j-2}.$$

Proof. Let $G \in \mathcal{M}(n,k,j)$. To prove the theorem by induction is it enough to show that maximal degree of G is at least $n+l$. (Because removing a vertex from G we obtain a graph in $\mathcal{M}(n,k-1,j)$ and clearly $m(n,j,j) = n$.)

Let $H = \{x \in V(G) : d(x) \geq n\}$ and $h = |H|$. Let us show that h is not large by applying the following procedure to G.

Let $A = C = \emptyset$ and $B = V(G)$. Repeat the following as long as possible or until $|A| = l+1$. Move to A any vertex $x \in B$ (if it exists) having at least n neighbours in B. For every such x do the *n-check*, that is, move to C all $y \in B \cap H$ with $d_{B\cup C}(y) < n$. (In fact, for every such y we have $d_{B\cup C}(y) = n-1$.)

Suppose we have stopped. Let a, b, c be the sizes of the eventual sets A, B, C. Inductively, we find a set $Y = \{y_1, \ldots, y_{l+1-a}\} \subset B \cup C$ such that each y_i has at least n neighbours in $C \cup B \setminus \{y_1, \ldots, y_{i-1}\}$. As each $y \in C$ has fewer than n neighbours in $B \cup C$, we conclude that $Y \subset B$. Let $R = (B \setminus Y) \cap H$ and $r = |C \cup R|$. Each $x \in R$ has at least n neighbours in $C \cup B$ for otherwise it would belong to C.

Counting the number of edges encountered in our algorithm we obtain that

$$e(G) \geq an + |Y|n + r(n-1) - \tbinom{r}{2} - r|Y|.$$

Using $a + |Y| = l+1$ (and the trivial inequality $|Y| \leq l$) we obtain

$$\tbinom{l+1}{2} \geq r\left(n - 1 - \tfrac{r-1}{2} - |Y|\right) \geq r\left(n - l - \tfrac{r+1}{2}\right).$$

To satisfy this quadratic in r inequality, r must not lie between the roots $r_{1,2} = n - l - \frac{1}{2} \pm R$, where

$$R = \frac{1}{2}\sqrt{4n^2 - 4n(2l+1) + 1}.$$

The assumption of the theorem implies that

$$l \leq 3n/8. \tag{112}$$

Using (112), one can check that $R \geq (n-3)/2$. Suppose that $r \geq r_2$. Observe that

$$r_2 \geq n - l - \tfrac{1}{2} + \tfrac{n-3}{2} = \tfrac{3}{2}n - l - 2 \geq \tfrac{9}{8}n - 2. \tag{113}$$

As before, the inequality $e(G) \geq an + |Y|n + r\frac{n-1}{2} - r|Y|$ implies that

$$\binom{l+1}{2} \geq r(\tfrac{n-1}{2} - l). \tag{114}$$

Using (112) and (113), we can deduce from (114) that $n \geq \frac{9}{128}n^2 + 1$, which cannot be satisfied for $n \geq 14$.

The above contradiction implies that $r \leq r_1$; then we have

$$h = r + l + 1 \leq n + \tfrac{1}{2} - R. \tag{115}$$

Suppose on the contrary that $k > j$ and $\Delta(G) < n + l$. For every $x \in H$ we choose a j-set $D_x \subset \overline{\Gamma(x)}$ and let $D = \cup_{x \in H} D_x$. We have $|D| \leq jh$ and we claim that this does not exceed $n + j$. To verify this, it is enough to check by (115) that $jR \geq jn - n - j/2$. Squaring, we obtain that the latter is equivalent to $n(2j-1) \geq 2j^2 l + j$, which is precisely our assumption.

Complete D to an arbitrary $(n+j)$-set E. As $G \in \mathcal{M}(n,k,j)$, some $x \in E \cap H$ has at least n neighbours in E, which is a contradiction as, by definition, E contains at least j non-neighbours of x.

Hence, $\Delta(G) \geq n + l$ and the theorem follows by induction. ∎

24 Splitting into Parts

Here we consider $b(n,m) = \min\{e(G) : G \in \mathcal{B}(n,m)\}$, where $\mathcal{B}(n,m)$ consists of all graphs G such that, for any partition $A \cup B = V(G)$, either $\Delta(G[A]) \geq n$ or $\Delta(G[B]) \geq m$.

Clearly, $b(n,m) = b(m,n)$. Let us assume $n \geq m$.

24.1 General Bounds

The following simple argument gives a very good general lower bound on $b(n,m)$.

Let $G \in \mathcal{B}(n,m)$ be any graph. Set initially $A = V(G)$ and $B = \emptyset$. As long as $|B| \leq m$, move to B any $x \in A$ with $d_A(x) \geq n$. (Such a vertex exists, because obviously $\Delta(G[B]) < m$.)

When we finish, $|B| = m + 1$. Swap the sets A and B each with the other. (So that now $|A| = m + 1$.) Next, consecutively and as long as possible, move to A any vertex of $G[B]$ of degree at least m. As eventually $\Delta(G[B]) < m$, our assumption on G implies that $|A| \geq n + 1$ (to allow a vertex of degree at least n). Counting the edges encountered in this procedure, we obtain the following bound valid for any n and m.

$$b(n, m) \geq (m + 1)n + ((n + 1) - (m + 1))m = 2mn - m^2 + n. \tag{116}$$

Next, we provide a general construction giving an upper bound on $b(n, m)$.

Example 86 *Choose representations $m = m_1 + \ldots + m_f$ and $n - m = n_1 + \ldots + n_g$. Let G be the disjoint union of $P_{m_i,n}$, $i \in [f]$, and $P_{n_j,m}$, $j \in [g]$, plus a vertex x connected to everything else. We claim that $G \in \mathcal{B}(n, m)$.*

Proof. Let $V(G) = A \cup B$ be any partition.

Case 1 Suppose $x \in A$. Observe that at least m_i vertices from each $P_{m_i,n}$ and at least n_j vertices from each $P_{n_j,m}$ lie in A. (Otherwise $\Delta(G[B]) \geq m$.) But then

$$d_A(x) = |A| - 1 \geq \sum_{i\in[f]} m_i + \sum_{j\in[g]} n_j = n.$$

Case 2 If $x \in B$ (and $\Delta(G[A]) < n$), then from each $P_{m_i,n}$ at least m_i vertices go to B and $d_B(x) \geq \sum_{i\in[f]} m_i = m$, as required. ∎

Let us compute how many edges we use in Example 86.

$$b(n, m) \leq e(G) = n+fn+gm+mn+(n-m)m+\sum_{i\in[f]} \binom{m_i}{2} + \sum_{j\in[g]} \binom{n_j}{2}, \tag{117}$$

To minimize it, we let the m_i's (and the n_j's) be nearly equal while f and g have to be around $m(2n)^{-1/2}$ and $(n-m)(2m)^{-1/2}$ respectively. Putting bounds (116) and (117) together we obtain the equality (108) claimed in the introduction.

24.2 Small Fixed m

In the extreme case when m is fixed and n tends to infinity, we consider Example 86 with $f = 1$ (so $m_1 = m$) and $g = n(2m)^{-1/2} + O(1)$. Then

$$\sum_{j\in[g]} \binom{n_j}{2} < g\frac{\left(\frac{n}{g} + 1\right)\frac{n}{g}}{2} \leq \frac{n}{2}(\sqrt{2m} + 1) + O(1),$$

and we obtain the following.

Corollary 87 *Let $m \geq 1$ be fixed. Then $b(n,m)$, $n \in \mathbb{N}$, lie between two linear functions, namely*

$$(2m+1)n + O(1) \leq b(n,m) \leq (2m + \sqrt{2m} + \tfrac{5}{2})n + O(1). \blacksquare$$

However, for a few particular small instances of m we can be more precise.

Let us provide a construction of $G \in \mathcal{B}(n,1)$, $n \geq 2$. Represent $n = 2k+l+1$ and let G be disjoint union of k triangles, l disjoint edges, plus vertices x, y; x is connected to every other vertex while y is connected to some n vertices (besides x). Clearly, $e(G) = 3k + 3k + l + 2l + n + 1 = 4n - 2$.

To show that $G \in \mathcal{B}(n,1)$, suppose that we have a partition $V(G) = A \cup B$ with B being an independent set. If one of x or y belongs to B, then A contains the other plus their n common neighbours and so $\Delta(G[A]) \geq n$. If $\{x,y\} \subset A$, then at least 2 vertices from each triangle and at least 1 vertex from each edge must be in A and $d_A(x) \geq 1 + 2k + l = n$, as required.

Theorem 88 *For any $n \geq 8$, $b(n,1) = 4n-2$ and all extremal graphs are given by the above construction.*

Proof. Let $n \geq 1$ and let G be any graph in $\mathcal{B}(n,1)$ of size at most $4n-2$. Let L be the set of vertices of G of degree at least n. Clearly, $|L| > 1$.

First, suppose that $|L| = 3$, say $L = \{x,y,z\}$. The partition with $B = L$ shows that L is not independent in G.

Case 1 $G[L]$ consists of a single edge, say $\{x,y\}$. The partition with $B = \{x,z\}$ (resp. $B = \{y,z\}$)) shows that y (resp. x) has at least n neighbours outside L. Hence, L is incident to at least $3n+1$ edges, and we have at most $(4n-2) - (3n+1) = n-3$ edges non-incident to L. Letting A consist of all vertices of L plus an arbitrary endvertex of each edge outside L, we obtain a contradiction: $\overline{A}$ is independent while $|A| \leq n$.

Case 2 $G[L]$ consists of two edges, say $\{x,y\}$ and $\{x,z\}$. The partition with $B = \{y,z\}$ shows $d_{\overline{L}}(x) \geq n$. The partition with $B = \{x\}$ shows that another vertex of L has at least n neighbours outside L. Hence, L is incident to at least $3n+1$ edges, and we can derive a contradiction as above.

Case 3 $G[L]$ is the complete graph. Placing in B a vertex of L, we deduce that some two vertices in L, say x and y, have at least $n+1$ neighbours each. Thus, we have already found $3n-1$ edges incident to L; so we can have at most one more such edge.

Case 3.1 Suppose that $d(z) = n$ and $d(x) \geq d(y)$. (That is, $d(y) = n+1$.) Every neighbour u of y is connected to x. (Otherwise consider $B = \{x, u\}$.) This means that $|U| = n-1$, where $U = (\Gamma(x) \cup \Gamma(y)) \setminus L$. Choose any $u \in U \setminus \Gamma(z) \neq \emptyset$. The partition with $B = \{z, u\}$ shows that $d(x) = n+2$. By letting $A = L$ and consecutively moving to A a non-isolated vertex of $G[\overline{A}]$, we conclude that $G[\overline{L}]$ consists of $n-2$ disjoint edges. (And $e(G) = 4n-2$.) Furthermore, if $n \geq 5$, we can choose an independent 3-set $C \subset \Gamma_{\overline{L}}(x)$. If z sends at least one edge to C, let $B = C$; otherwise let $B = C \cup \{z\}$. It is easy to see that in either case $\Delta(G[\overline{B}]) < n$, which is a contradiction.

Case 3.2 Suppose $d(x) = d(y) = d(z) = n+1$. As before, we conclude that $G[\overline{L}]$ consists of $n-2$ disjoint edges. (And $e(G) = 4n-2$.) Let $n \geq 7$. Clearly, $|V| \leq 2$, where $V = \Gamma(x) \cap \Gamma(y) \cap \Gamma(z)$. (Otherwise, consider any independent 2-set $B \subset L$.) Also, there is no $v \in (\Gamma(x) \cap \Gamma(y)) \setminus \Gamma(z)$. (Otherwise, let $B = \{v, z\}$.) But then, for $n \geq 5$, we can choose non-incident $u, v \in \overline{\Gamma(z)}$ with $u \in \Gamma(x)$ and $v \in \Gamma(y)$, and the consideration of $B = \{u, v, z\}$ yields a contradiction.

Similarly, but with less effort, we can exclude the case $|L| \geq 4$ for $n \geq 6$. So, we conclude that $L = \{x, y\}$. Considering the partition with $B = \{x, y\}$, we see that x is connected to y. Considering the partition with $B = \{x\}$ or $B = \{y\}$, we conclude that $d(x) \geq n+1$ and $d(y) \geq n+1$.

We apply the following procedure. Let A consist of x and y plus all vertices connected neither to x nor to y; let $B = \overline{A}$. At Stage 0 consecutively move to A a vertex of degree at least 3 in $G[B]$. Stage 1: one by one and as long as possible, move to A a vertex of degree 2 in $G[B]$.

Now, $G[B]$ consists of isolated edges. Stage 2: for each edge $\{a, b\} \in E(G[B])$ with $\Gamma(a) \cap L \subset \Gamma(b) \cap L$ we move a to A but keep b in B.

After this stage each edge in $G[B]$ together with L spans a C_4; let $s_3 = e(G[B])$. For $j \in [0, 2]$ let $s_{j,i}$ be the number of vertices moved to A at Stage j which were incident to i vertices in L, $i \in [2]$, and let $s_j = s_{j,1} + s_{j,2}$.

Case 4 Suppose that $\Delta(G[A]) \geq n$; let $d_A(x) \geq n$. Then the total number of edges in G is at least $n+1$ (edges incident to y) plus $3(n-1)$. (Because for each of $n-1$ vertices incident to x which were moved to A we count at least three edges; for example, for a vertex a moved at Stage 2, we encounter the edges $\{a,x\}$, $\{a,b\} \in E(G[B])$ and $\{b,x\}$.) Hence, $e(G) \geq 4n-2$ as required.

Case 5 Suppose that $\Delta(G[A]) < n$. As we can make $B \subset V(G)$ independent by moving an arbitrary endvertex of each edge to A (and after this we must have $\Delta(G[A]) \geq n$), we conclude that now

$$2n-1 \leq s_3 + d_A(x) + d_A(y) = s_3 + 2 + \sum_{j=0}^{2}(s_{j,1} + 2s_{j,2}). \tag{118}$$

On the other hand, we have the following estimate.

$$e(G) \geq 4s_{0,1} + 5s_{0,2} + 3s_{1,1} + 4s_{1,2} + \max(s_1 - s_2 - s_3, 0) + 3s_{2,1} + 5s_{2,2} + 3s_3 + 1. \tag{119}$$

Only the max-term needs some explanation. After Stage 1 $G[B]$ consists of $s_3 + s_4$ isolated edges. Let us move back to B the s_1 vertices moved at Stage 1. As the resulting graph has maximal degree 2, we must use at least $s_1 - s_2 - s_3$ new vertices. Each of these vertices sends at least one edge to L, which constitutes the extra term. If we multiply (118) by 2 and substitute this from (119), we obtain (using $e(G) \leq 4n-2$)

$$s_3 + 2s_{0,1} + s_{0,2} + s_{1,1} + s_2 + \max(s_1 - s_2 - s_3, 0) \leq 3.$$

Hence, $s_1 \leq 3$ (and $s_2 + s_3 \leq 3$). From (118) we deduce that

$$n \leq (s_1 + s_2) + (s_3 + 3)/2 \leq 7\tfrac{1}{2}.$$

Hence, we have shown that $b(n,1) = 4n-2$, for $n \geq 8$. Conversely, if a graph G achieves this bound, then $|L| = 2$, $s_0 = 0$, $s_3 = 0$, one vertex of L is connected to every $\overline{L}$-neighbour of the other L-vertex, and every vertex moved at Stage 1 belonged to an isolated triangle of $G[B]$. Now the required characterization follows. The details are left to the reader. ▮

Remark. Perhaps, $b(n,1) = 4n-2$ for any $n \geq 2$ (clearly $b(1,1) = 3$), but then there are many other constructions achieving this bound. A direct search is feasible (note that our proof of Theorem 88 contains some information reducing this search), but it would be too long to include here.

Theorem 89 $b(n,2) = 6n + O(1)$.

Proof. A construction of $G \in \mathcal{B}(n,2)$ first. Consider $\lceil \frac{n}{2} \rceil$ disjoint 4-cycles and one triangle, say on $X = \{x_1, x_2, x_3\}$. To this we add some further edges: x_1 is connected to every other vertex while x_2 and x_3 are connected to some fixed m-set $C \subset \Gamma(x_1) \setminus \{x_2, x_3\}$.

Let $V(G) = A \cup B$ be any partition with $\Delta(G[B]) \leq 1$. If $x_1 \in A$, then at least 2 vertices of each C_4 belong to A and $d_A(x_1) = |A| - 1 \geq n$, as required. If $x_1 \in B$, then all but at most one vertex in $C \cup \{x_2, x_3\}$ lie in A and a vertex in $X \cap A \neq \emptyset$ has at least $|C| = n$ neighbours in A. Hence, $G \in \mathcal{B}(n,2)$ and $b(n,2) \leq 6n + O(1)$, as required.

We show the lower bound. Let $G \in \mathcal{B}(n,2)$ be any graph with at most $6n$ edges; we have to deduce $6n - e(G) = O(1)$. Let $L = \{x \in V(G) : d(x) \geq n\}$.

If $|L| \geq 4$ then we have at least $4n + O(1)$ edges incident to L. Let $A = L$ and $B = \overline{A}$. As long as possible, move to A a vertex of $G[B]$ of degree at least 2. Before we stop, we repeat the iteration at least $n + 1 - |L| = n + O(1)$ times, which means that there are $2n + O(1)$ edges not incident to L and we are home.

Clearly, $|L| \geq 2$. (Otherwise the partition with $B = L$ contradicts $G \in \mathcal{B}(n,2)$.) Hence, $|L| = 3$ and the theorem follows from Lemma 90 below. ∎

The following related notion is useful. Let $\mathcal{B}'(n,m,l)$ be the class of graphs G with a fixed l-set $L \subset V(G)$ such that $d(x) \geq n$, $x \in L$, and for any partition $V(G) = A \cup B$ with $\Delta(G[B]) < m$ and $L \subset A$ some vertex $x \in L$ has at least n neighbours in A. Also, denote $b'(n,m,l) = \min\{e(G) : G \in \mathcal{B}'(n,m,l)\}$.

Lemma 90 *For* $l \in [3]$, $b'(n,2,l) \geq (3+l)n + O(1)$.

Proof. Let $G \in \mathcal{B}'(n,2,l)$ be any graph. We may freely remove any vertex incident to no vertex of the selected set $L = \{x_1, \ldots, x_l\}$, as this does not violates the $\mathcal{B}'(n,2,l)$-property. Let

$$\Gamma_A = \left\{ y \in \overline{L} : \{y, x_i\} \in E(G) \text{ iff } i \in A \right\}, \quad A \subset [l].$$

Case 1 Let $l = 1$. Let $A = L$ and $B = V(G) \setminus L$. As long as possible, move to A any $y \in B$ with $d_B(y) \geq 3$. At the end, $G[B]$ consists of disjoint cycles, paths and vertices. But we can move to A at most $\lfloor \frac{p+2}{3} \rfloor$ (resp. $\lfloor \frac{p+1}{3} \rfloor$) vertices from each cycle (resp. path) of length p to ensure $\Delta(G[B]) < 2$. As the number of moved vertices must be at least n and we use at least 4 edges per vertex (including edges incident to x_1), the claim follows.

Case 2 Let $l = 2$. We apply an inductive on n argument, ensuring that we have at least 5 edges per every removed vertex, except in $O(1)$ cases. First, whenever we have $y \in \overline{L}$ with $d_G(y) \geq 5$, we remove it, obtaining a graph in $\mathcal{B}'(n-1,2,2)$. Next, if we have $y_1 \in \Gamma_1$ and $y_2 \in \Gamma_2$ at distance at least 3, we contract them without loosing the $\mathcal{B}'(n,2,2)$-property. Suppose we are finally stuck and suppose $|\Gamma_1| \leq |\Gamma_2|$. As $\Delta(G[\overline{L}]) < 4$, we conclude that $g = |\Gamma_1| = O(1)$. Removing Γ_1 from G, we obtain a graph in $\mathcal{B}'(n-g,2,2)$; further, removing x_2 (and at least $n-g$ edges) we obtain a graph in $\mathcal{B}'(n-g-1,2,1)$ which has size at least $4n + O(1)$. Hence, $b'(n,2,2) \geq 5n + O(1)$.

Case 3 Let $m = 3$. Like in Case 2, we remove a vertex $x \in \overline{L}$ of degree at least 6; also, we contract any $y \in \Gamma_A$, $z \in \Gamma_B$, at distance at least 3 for $A \cap B = \emptyset$. Next, removing $O(1)$ vertices we ensure that all but one of Γ_i, $i \in [3]$, are empty, say $\Gamma_1 = \Gamma_2 = \emptyset$. Also, we make either Γ_{12} or Γ_3 empty. If $\Gamma_{12} = \emptyset$, then $\Gamma(x_1) \subset \Gamma(x_3)$; removing x_1 (and $\geq n + O(1)$ edges) we obtain a graph in $\mathcal{B}'(n+O(1),2,2)$ of size at least $5n + O(1)$—we are home.

So, suppose $\Gamma_3 = \emptyset$. If possible, remove any three vertices in respectively $\Gamma_{12}, \Gamma_{13}, \Gamma_{23}$ incident to at least 12 edges to obtain a graph in $\mathcal{B}'(n-2,2,3)$. Removing up to $O(1)$ vertices, we can assume that $d_{\overline{L}}(y) = 1$ for each y in, for example, Γ_{12}. Let $z \in \overline{L}$ be the neighbour of some $y \in \Gamma_{12}$. If $d_{\overline{L}}(z) = 1$, then we can remove y, z from G without violating the $\mathcal{B}'(n,2,3)$-property; otherwise, removing y, z we remove at least 6 edges and obtain a $\mathcal{B}'(n-1,2,3)$-graph. Eventually, we achieve $\Gamma_{12} = \emptyset$, that is, $\Gamma(x_1) \subset \Gamma(x_3)$ and we are home again by Case 2. ∎

Remark. In the next case $m = 3$ we can only show that

$$7n + O(1) \leq b(n,3) \leq 9n + O(1).$$

References

[AEHK96] N. Alon, P. Erdős, R. Holzman, and M. Krivelevich. On k-saturated graphs with restrictions on the degrees. *J. Graph Theory*, 23:1–20, 1996.

[AF86] R.P. Anstee and Z. Füredi. Forbidden submatrices. *Discrete Math.*, 62:225–243, 1986.

[AGS97] R. P. Anstee, J. R. Griggs, and A. Sali. Small forbidden configurations. *Graphs Combin.*, 13:97–118, 1997.

[Alo85] N. Alon. An extremal problem for sets with applications to graph theory. *J. Combin. Theory* (A), 40:82–89, 1985.

[And67] I. Anderson. *Some Problems in Combinatorial Number Theory.* PhD thesis, Univ. of Nottingham, 1967.

[And87] I. Anderson. *Combinatorics of Finite Sets.* Oxford Univ. Press, 1987.

[Ans95] R. P. Anstee. Forbidden configurations: Induction and linear algebra. *Europ. J. Combin.*, 16:427–438, 1995.

[BCE$^+$96] C. A. Barefoot, L. H. Clark, R. C. Entringer, T. D. Porter, L. A. Székely, and Z. Tuza. Cycle-saturated graphs of minimum size. *Discrete Math.*, 150:31–48, 1996.

[Bei68] L. W. Beineke. On derived graphs and digraphs. In *Beitr. Graphentheorie, Int. Kolloquium Manebach (DDR)*, pages 17–23, 1968.

[Blo87] A. Blokhuis. More on maximal intersecting families of finite sets. *J. Combin. Theory* (A), 44:299–303, 1987.

[BMRR80] L. D. Baumert, R. J. McEliece, R. J. Rodermich, and H. Rumsey. A probabilistic version of Sperner's theorem. *Ars Combinatoria*, 9:91–100, 1980.

[Bol65] B. Bollobás. On generalized graphs. *Acta Math. Acad. Sci. Hung.*, 16:447–452, 1965.

[Bol67a] B. Bollobás. Determination of extremal graphs by using weights. *Wiss. Z. Hochsch. Ilmenau*, 13:419–421, 1967.

[Bol67b] B. Bollobás. On a conjecture of Erdős, Hajnal and Moon. *Amer. Math. Monthly*, 74:178–179, 1967.

[Bol67c] B. Bollobás. Weakly k-saturated graphs. In *Proc. Coll. Graph Theory*, pages 25–31. Ilmenau, 1967.

[Bol74] B. Bollobás. Three-graphs without two triples whose symmetric difference is contained in a third. *Discrete Math.*, 8:21–24, 1974.

[Bol76] B. Bollobás. On complete subgraphs of different orders. *Math. Proc. Camb. Phil. Soc.*, 79:19–24, 1976.

[Bol78] B. Bollobás. *Extremal Graph Theory.* Academic Press, London, 1978.

[Bol86] B. Bollobás. *Combinatorics, Set Systems, Families of Vectors, and Combinatorial Probability.* Cambridge Univ. Press, 1986.

[Bol95] B. Bollobás. Extremal graph theory. In R. L. Graham, M. Grötschel, and L. Lovász, editors, *Handbook of Combinatorics*, pages 1231–1292. Elsevier Science B.V., 1995.

[Bon72a] J. A. Bondy. Induced subsets. *J. Combin. Theory* (B), 12:201–202, 1972.

[Bon72b] J. A. Bondy. Variations on the hamiltonian scheme. *Can. Math. Bull.*, 15:57–62, 1972.

[Bou74] N. Bourbaki. *Algebra I.* Reading: Addison-Wesley, 1974.

[BP69] L. W. Beineke and R. E. Pippert. The number of labelled k-dimensional trees. *J. Combin. Theory*, 6:200–205, 1969.

[BP77] L. W. Beineke and R. E. Pippert. On the structure of (m,n)-trees. In *Proc. 8th Southeast Conference on Combinatorics, Graph Th. and Computing (Louisiana State Univ., Baton Rouge)*, volume XIX of *Congressus Numeratium*, pages 75–80. Utilitas Math., Winnipeg, 1977.

[BS74] J. A. Bondy and M. Simonovits. Cycles of even length in graphs. *J. Combin. Theory* (B), 16:97–105, 1974.

[BTK51] N. G. de Bruijn, C. Tengbergen, and D. Kruyswijk. On the set of divisors of a number. *Nieuw Arch. Wiskd.*, 23:191–193, 1951.

[Cam95] P. J. Cameron. Counting two-graphs related to trees. *Electronic J. Combin.*, 2(#R4):8pp, 1995.

[Cay89] A. Cayley. A theorem on trees. *Quart. J. Math.*, 23:376–378, 1889.

[CCES86] L. H. Clark, R. P. Crane, R. C. Entringer, and H. D. Shapiro. On smallest maximally non-hamiltonian graphs. In *Combinatorics, Graph Theory and Computing, Proc. 17th Southeast. Conf., Boca Raton*, volume 53 of *Congres. Numer.*, pages 215–220, 1986.

[CE83] L. H. Clark and R. C. Entringer. Smallest maximally non-hamiltonian graphs. *Period. Math. Hungar*, 14:57–68, 1983.

[CES92] L. H. Clark, R. C. Entringer, and H. D. Shapiro. Smallest maximally nonhamiltonian graphs II. *Graphs Combin.*, 8:225–231, 1992.

[Che52] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statistics*, 23:493–507, 1952.

[Che93] W. Y. C. Chen. A coding algorithm for Rényi trees. *J. Combin. Theory* (A), 63:11–25, 1993.

[Chu97] F. R. K. Chung. Open problems of Paul Erdős in graph theory. *J. Graph Theory*, 25:3–36, 1997.

[CL85] C.-Y. Chao and N.-Z. Li. Trees of polygons. *Archiv Math.*, 45:180–185, 1985.

[CO82] F. Chin and G. Ozsoyoglu. Auditing and inference control in statistical databases. *IEEE Trans. Software Eng.*, SE–8:574–582, 1982.

[dCKW88] D. de Caen, D. L. Kreher, and J. Wiseman. On constructive upper bounds for the Turán numbers $T(n, 2r + 1, 2r)$. In *Combinatorics, Graph Theory and Computing, Proc. 19th Southeast. Conf., Baton Rouge*, volume 65 of *Congres. Numer.*, pages 277–280, 1988.

[DDFL85] S. J. Dow, D. A. Drake, Z. Füredi, and J. A. Larson. A lower bound for the cardinality of a maximal family of mutually intersecting sets of equal size. In *Combinatorics, Graph Theory and Computing, Proc. 16th Southeast. Conf., Boca Raton*, volume 48 of *Congres. Numer.*, pages 47–48, 1985.

[Dew74] A. K. Dewdney. Multidimensional tree-like structures. *J. Combin. Theory* (B), 17:160–169, 1974.

[DH56] G. B. Dantzig and A. J. Hoffman. Dilworth's theorem on partially ordered sets. In H. W. Kuhn and A. W. Tucker, editors, *Linear Inequalities and Related Systems*, volume 38 of *Annals Math. Studies*, pages 207–214. Princeton Univ. Press, 1956.

[DH86] D. A. Duffus and D. Hanson. Minimal k-saturated and color critical graphs of prescribed minimum degree. *J. Graph Theory*, 10:55–67, 1986.

[Dil50] R. P. Dilworth. A decomposition theorem for partially ordered sets. *Annals of Math.*, 51(2):161–166, 1950.

[Doo73] M. Doob. An interrelation between line graphs, eigenvalues, and matroids. *J. Combin. Theory* (B), 15:40–50, 1973.

[EF99] P. Erdős and R. Faudree. Restricted size Ramsey numbers for cycles and stars. To appear in *Discrete Math.*, 1999.

[EFT91] P. Erdős, Z. Füredi, and Z. Tuza. Saturated r-uniform hypergraphs. *Discrete Math.*, 98:95–104, 1991.

[EG59] P. Erdős and T. Gallai. On maximal paths and circuits of graphs. *Acta Math. Acad. Sci. Hung.*, 10:337–356, 1959.

[EG61] P. Erdős and T. Gallai. On the minimal number of vertices representing the edges of a graph. *Publ. Math. Inst. Hungar. Acad. Sci.*, 6:181–203, 1961.

[EH94] P. Erdős and R. Holzman. On maximal triangle-free graphs. *J. Graph Theory*, 18:585–594, 1994.

[EHM64] P. Erdős, A. Hajnal, and J. W. Moon. A problem in graph theory. *Amer. Math. Monthly*, 71:1107–1110, 1964.

[Eng97] K. Engel. *Sperner Theory.* Cambridge Univ. Press, 1997.

[ER66] J. Edmonds and G.-C. Rota. Submodular set functions (Abstract). *Waterloo Combinatorics Conference*, 1966.

[Erd81] P. Erdős. Problems and results in graph theory. In G. Chartrand, editor, *The Theory and Applications of Graphs*, pages 331–341. John Wiley, New York, 1981.

[Erd99] P. Erdős. A selection of problems and results in combinatorics. *Combin. Prob. Computing*, 8:1–6, 1999.

[ERSS96] P. Erdős, T. J. Reid, R. Schelp, and W. Staton. Sizes of graphs with induced subgraphs of large maximum degree. *Discrete Math.*, 158:283–286, 1996.

[ES83] P. Erdős and M. Simonovits. Supersaturated graphs and hypergraphs. *Combinatorica*, 3:181–192, 1983.

[ES88] O. Egecioglu and L.-P. Shen. A bijective proof for the number of labelled q-trees. *Ars Combinatoria*, 25:3–30, 1988.

[FFP87] P. Frankl, Z. Füredi, and J. Pach. Bounding one-way differences. *Graphs Combin.*, 3:341–347, 1987.

[FHPZ98] Z. Füredi, P. Horak, C. M. Pareek, and X. Zhu. Minimal oriented graphs of diameter 2. *Graphs Combin.*, 14:345–350, 1998.

[Foa71] D. Foata. Enumerating k-trees. *Discrete Math.*, 1:181–186, 1971.

[Fra82] P. Frankl. An extremal problem for two families of sets. *Europ. J. Combin.*, 3:125–127, 1982.

[Fra95] P. Frankl. Extremal set systems. In R. L. Graham, M. Grötschel, and L. Lovász, editors, *Handbook of Combinatorics*, pages 1293–1330. Elsevier Science B.V., 1995.

[FRS97] R. J. Faudree, C. C. Rousseau, and R. H. Schelp. Problems in graph theory from Memphis. In R. L. Graham, editor, *The Mathematics of Paul Erdős*, volume 2, pages 7–26. Springer, Berlin, 1997.

[FS94] Z. Füredi and A. Seress. Maximal triangle-free graphs with restrictions on degrees. *J. Graph Theory*, 18:11–24, 1994.

[Gho75] S. P. Ghosh. Consecutive storage of relevant records with redundancy. *Comm. ACM*, 18:464–471, 1975.

[GI75] C. Greene and G. A. Iba. Caley's formula for multidimensional trees. *Discrete Math.*, 14:1–11, 1975.

[GK76] C. Greene and D. J. Kleitman. Strong versions of Sperner's theorem. *J. Combin. Theory* (A), 20:80–88, 1976.

[GKP89] R. L. Graham, D. E. Knuth, and O. Patashnik. *Concrete Mathematics: a Foundation for Computer Science.* Addison-Wesley Publ. Comp., 1989.

[GKP95] D. M. Gordon, G. Kuperberg, and O. Patashnik. New constructions for covering designs. *J. Combin. Designs*, 3:269–284, 1995.

[GPKS96] D. M. Gordon, O. Patashnik, G. Kuperberg, and J. H. Spencer. Asymptotically optimal covering designs. *J. Combin. Theory* (A), 75:270–280, 1996.

[Gri77] J. R. Griggs. Sufficient conditions for a symmetric chain order. *SIAM J. Appl. Math.*, 32:807–809, 1977.

[GST98] G. R. Griggs, M. Simonovits, and G. R. Thomas. Extremal graphs with bounded densities of small subgraphs. *J. Graph Theory*, 29:185–207, 1998.

[Haj65] A. Hajnal. A theorem on k-saturated graphs. *Can. J. Math.*, 17:720–724, 1965.

[Han66] G. Hansel. Sur de nombre des fonctions Booléennes monotones de n variables. *C. R. Acad. Sci. Paris Sér. A-B*, 262:1088–1090, 1966.

[HS84] D. Hanson and K. Seyffarth. k-saturated graphs of prescribed maximum degree. *Congres. Numer.*, 42:169–182, 1984.

[HT87] D. Hanson and B. Toft. Edge-colored saturated graphs. *J. Graph Theory*, 11:191–196, 1987.

[HT91] D. Hanson and B. Toft. k-saturated graphs of chromatic number at least k. *Ars Combinatoria*, 31:159–164, 1991.

[Isa75] R. Isaacs. Infinite families of non-trivial trivalent graphs which are not Tait colorable. *Amer. Math. Monthly*, 82:221–239, 1975.

[Kal84] G. Kalai. Weakly saturated graphs are rigid. *Ann. Discrete Math.*, 20:189–190, 1984.

[Kal85] G. Kalai. Hyperconnectivity of graphs. *Graphs Combin.*, 1:65–79, 1985.

[Kal90] G. Kalai. Symmetric matroids. *J. Combin. Theory* (B), 50:54–64, 1990.

[Kat66] 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.

[Kat74] G. O. H. Katona. Extremal problems for hypergraphs. In *Combinatorics*, volume 56, pages 13–42. Math. Cent. Tracts, 1974.

[KNS64] G. O. H. Katona, T. Nemetz, and M. Simonovits. On a graph problem of Turán (In Hungarian). *Mat. Fiz. Lapok*, 15:228–238, 1964.

[Kor81] A. D. Korshunov. The number of monotone Boolean functions. *Probl. Kibernet.*, 38:5–108, 1981.

[Kos82] A. V. Kostochka. A class of constructions for Turán $(3,4)$-problem. *Combinatorica*, 2:187–192, 1982.

[Kru63] 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.

[KS67] G. O. H. Katona and E. Szemerédi. On a problem of graph theory. *Studia Sci. Mat. Hung.*, 2:23–28, 1967.

[KS79] D. J. Kleitman and J. B. Shearer. Probabilities of independent choices being ordered. *Stud. Appl. Math.*, 60:271–276, 1979.

[KT86] L. Kászonyi and Z. Tuza. Saturated graphs with minimal number of edges. *J. Graph Theory*, 10:203–210, 1986.

[KT96] K. M. Koh and C. P. Teo. Chromaticity of series-parallel graphs. *Discrete Math.*, 154:289–295, 1996.

[Lip78] W. Lipski. On strings containing all subsets as substrings. *Discrete Math.*, 21:253–259, 1978.

[Lov77] L. Lovász. Flats in matroids and geometric graphs. In P. J. Cameron, editor, *Proc. 6th British Combin. Conf.*, pages 45–86. Academic Press, 1977.

[Mar75] M. Marcus. *Finite Dimensional Multilinear Algebra, I, II.* Marcel Dekker, New York, 1973, 1975.

[Moo69] J. W. Moon. The number of labeled k-trees. *J. Combin. Theory*, 6:196–199, 1969.

[MRS91] M. Miller, I. Roberts, and J. Simpson. Application of symmetric chains to an optimization problem in the security of statistical databases. *Bull. ICA*, 2:47–58, 1991.

[MS95] K. C. Miller and D. G. Sarvate. An application of symmetric chains to a statistical database compromise prevention problem. *Bull. ICA*, 13:57–64, 1995.

[Oll72] L. T. Ollman. $K_{2,2}$-saturated graphs with a minimal number of edges. In *Combinatorics, Graph Theory and Computing, Proc. 13th Southeast. Conf., Boca Raton,* (Utilitas Math., Winnipeg), pages 367–392, 1972.

[Oxl92] J. G. Oxley. *Matroid Theory.* Oxford Univ. Press, 1992.

[Pen93] Y.-H. Peng. On the chromatic uniqueness of certain trees of polygons. *J. Austral. Math. Soc. Ser. A*, 55:403–410, 1993.

[Pik97] O. Pikhurko. On edge decomposition of posets. Submitted to *Order*, 1997.

[Pik98] O. Pikhurko. Uniform families and count matroids. Submitted to *Graphs Comb.*, 1998.

[Pik99a] O. Pikhurko. Application of gross matroids to w-sat-problems. In preparation, 1999.

[Pik99b] O. Pikhurko. Asymptotic evaluation of the sat-function for r-stars. To appear in *Discrete Math.*, 1999.

[Pik99c] O. Pikhurko. Enumeration of labelled (k, m)-trees. *J. Combin. Theory* (A), 86:197–199, 1999.

[Pik99d] O. Pikhurko. The minimal size of saturated hypergraphs. To appear in *Combin. Prob. Comput.*, 1999.

[Pik99e] O. Pikhurko. On the triangle-vs-star size Ramsey number. Submitted to *Combinatorica*, 1999.

[PW70] M. J. Piff and D. J. A. Welsh. On the vector representation of matroids. *J. Lond. Math. Soc.*, 2:284–288, 1970.

[Röd85] V. Rödl. On a packing and covering problem. *Europ. J. Combin.*, 5:69–78, 1985.

[RR70] C. Rényi and A. Rényi. The Prüfer code for k-trees. In P. Erdős et al., editors, *Combinatorial Theory and Its Applications*, pages 945–971. North-Holland, Amsterdam, 1970.

[Sau73] N. Sauer. On the density of families of sets. *J. Combin. Theory* (A), 13:145–147, 1973.

[She72] S. Shelah. A combinatorial problem: Stability and order for models and theories in infinitary languages. *Pac. J. Math*, 4:247–261, 1972.

[Sid87] A. F. Sidorenko. Solution of a problem of Bollobás on 4-graphs. *Mat. Zametki*, 41:433–455, 1987.

[Sid95] A. Sidorenko. What we know and what we do not know about Turán numbers. *Graphs Combin.*, 11:179–199, 1995.

[ST98] R. H. Schelp and A. G. Thomason. A remark on the number of complete and empty subgraphs. *Combin. Prob. Computing*, 7:217–220, 1998.

[TT91] M. Truszczynski and Z. Tuza. Asymptotics results on saturated graphs. *Discrete Math.*, 87:309–314, 1991.

[Tuz86] Z. Tuza. A generalization of saturated graphs for finite languages. In *Proc. IMYCS'86, Smolenice Castle (Czechoslovakia)*, volume 185 of *MTA SZTAKI Studies*, pages 287–293, 1986.

[Tuz88] Z. Tuza. Extremal problems on saturated graphs and hypergraphs. *Ars Combinatoria*, 25B:105–113, 1988.

[Tuz89] Z. Tuza. C_4-saturated graphs of minimum size. *Acta Univ. Carolin. Math. Phys.*, 30:161–167, 1989.

[Tuz92] Z. Tuza. Asymptotic growth of sparse saturated structures is locally determined. *Discrete Math.*, 108:397–402, 1992.

[VC71] V. N. Vapnik and A. Chervonenkis. The uniform convergence of frequences of the appearance of events to their probabilities (in Russian). *Teor. Veroyatn. Primen.*, 16:264–279, 1971.

[Ver99] J. Verstraëte. On arithmetic progressions of cycle lengths in graphs. Preprint, 1999.

[Wel76] D. J. A. Welsh. *Matroid Theory*, volume 8 of *London Math. Soc. Monographs*. Academic Press, 1976.

[Wes66] W. Wessel. Über eine Klasse paarer Graphen, I: Beweis einer Vermutung von Erdős, Hajnal and Moon. *Wiss. Z. Hochsch. Ilmenau*, 12:253–256, 1966.

[Wes67] W. Wessel. Über eine Klasse paarer Graphen, II: Bestimmung der Minimalgraphen. *Wiss. Z. Hochsch. Ilmenau*, 13:423–426, 1967.

[Whi88] E. G. Whitehead, Jr. Chromatic polynomials of generalized trees. *Discrete Math.*, 72:391–393, 1988.

[Whi89] W. Whiteley. A matroid on hypergraphs, with applications in scene analysis and geometry. *Discrete Comput. Geom.*, 4:75–95, 1989.

[Whi96] 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.

[WW84] 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.

[XWCY97] L. Xiaohui, J. Wenzhou, Z. Chengxue, and Y. Yuansheng. On smallest maximally nonhamiltonian graphs. *Ars Combinatoria*, 45:263–270, 1997.

[Yu93] J. Yu. An extremal problem for sets: A new approach via Bezoutians. *J. Combin. Theory* (A), 62:170–175, 1993.

[Zyk49] A. A. Zykov. On some properties of linear complexes (in Russian). *Mat. Sbornik N.S.*, 24:163–188, 1949.

Index