Combinatorics, Probability and Computing (2008) 17, 641–660. © 2008 Cambridge University Press
doi:10.1017/S096354830800922X Printed in the United Kingdom

Generating All Sets With Bounded Unions

YANNICK FREIN, BENJAMIN LÉVÊQUE and ANDRÁS SEBŐ †

Laboratoire G-SCOP, INPG, UJF, CNRS,
46, avenue Felix Viallet, 38031 Grenoble Cedex, France
(e-mail: yannick.frein.benjamin.leveque.andras.sebo@g-scop.inpg.fr)

Received 19 November 2007; revised 27 May 2008; first published online 8 July 2008

We consider the problem of minimizing the size of a family of sets $\mathcal{G}$ such that every subset of $\{1,\dots,n\}$ can be written as a disjoint union of at most k members of $\mathcal{G}$, where k and n are given numbers. This problem originates in a real-world application aiming at the diversity of industrial production. At the same time, the question of finding the minimum of $|\mathcal{G}|$ so that every subset of $\{1,\dots,n\}$ is the union of two sets in $\mathcal{G}$ was asked by Erdős and studied recently by Füredi and Katona without requiring the disjointness of the sets. A simple construction providing a feasible solution is conjectured to be optimal for this problem for all values of n and k and regardless of the disjointness requirement; we prove this conjecture in special cases including all (n,k) for which $n \leqslant 3k$ holds, and some individual values of n and k.

1. Introduction

The n-element set $\{1,\dots,n\}$ is denoted by $[n]$. For two positive integers n,k, a family $\mathcal{G}$ of subsets of $[n]$ is said to *k-generate* $X \subseteq [n]$ if X is the disjoint union of at most k members of $\mathcal{G}$. It k-generates the family $\mathcal{H} \subseteq \mathcal{P}([n])$ if it k-generates every $X \in \mathcal{H}$. It is called an (n,k)-generator if it generates the entire powerset $\mathcal{P}([n])$, that is, if every non-empty subset of $[n]$ can be obtained as a disjoint union of at most k members of $\mathcal{G}$. This work aims at determining the (n,k)-generators of minimum size. The *size* of a set is the number of its elements (synonymous with cardinality).

Sets of size 1 are called *singletons*. All the singletons $\{i\}$ $(i = 1,\dots,n)$ must be contained in any (n,k)-generator. We call an (n,k)-generator $\mathcal{G}$ of minimum size *optimal*, and introduce the notation $E(n,k) := |\mathcal{G}|$. A generator can be represented by a *hypergraph* (family of sets) where the vertices are the elements of $[n]$ and the hyperedges are the members of $\mathcal{G}$.

† This research was supported by the ADONET network of the European Community, which is a Marie Curie Training Network, and the Centre for Advanced Study at the Norwegian Academy of Science and Letters, Oslo.

https://doi.org/10.1017/S096354830800922X Published online by Cambridge University Press

As Zoltán Füredi reports, Paul Erdős [2] asked about the case $k = 2$ allowing the target-sets to be *not necessarily disjoint* unions of two members of $\mathcal{G}$. He conjectured that optimal generators consist of all the non-empty subsets of V_1 and V_2, where V_1, V_2 is a partition of $[n]$ into two almost equal parts. Since every subset of $[n]$ is the *disjoint* union of two sets in this generator, it is implicit in this conjecture that the optimum value does not depend on whether or not the two sets in the definition are required to be disjoint.

Erdős also considered the problem of generating only sets of size at most s, where s is a positive integer. Füredi and Katona investigated this latter problem in [3]. For $s \leqslant 2$ the problem is void, and for $s = 3$ the problem is equivalent to Turán's theorem [6]. For $s \leqslant 4$, $n \geqslant 8$ they establish that the cardinality of an optimal generator is $n + \binom{n}{2} - \lfloor \frac{4}{3} n \rfloor$. When $s \leqslant 4$ it clearly does not matter whether or not the two sets are required to be disjoint. (The same may be true for $s > 4$ see Section 2, but we cannot prove this.) For all $s > 4$ the problem is apparently open.

The same questions have been asked independently for optimizing the diversity of production in the car industry. To answer market requirements, many companies want to reduce the delay between the ordering and the delivery of a finished product, in the context of offering a large choice for the possible options of these products. The industrial problem that has to be faced is the following: determine the *semi-finished products* – each of which corresponds to a set of options – that must be stocked in order to be able to assemble any possible finished product in at most a given number of operations [1]. This latter constraint guarantees an assembly time that does not exceed a desired time of delivery. The aim is to minimize the size of the stock under this constraint. This is equivalent to finding an optimal (n, k)-generator, where n is the number of options, and k the maximum number of semi-finished products that can be assembled. From the viewpoint of industrial technology the *disjointness* constraint cannot be relaxed, and it is better to be able to generate *all subsets*. Refining these constraints, the optimization problems that can be stated appear to be too difficult (NP-hard – see Section 5); on the other hand, these rigid requirements bring us to the prefixed constraints of extremal combinatorics versus the flexible inputs of algorithmic problems. These questions lead directly to beautiful and seemingly difficult mathematical problems.

The basic problem studied in this article has been mentioned by the first author in the activity report of the project 'Decision Making Under Uncertainty' at the Centre for Advanced Study of Oslo, in 2000–2001. Conjecture 2.1 below is explicitly mentioned in [1] independently of Erdős [2]. However, the only result on this problem so far seems to be [3].

In Section 2 we introduce the main construction and provide the related conjectures, remarks and some other preliminaries, including the relation of the problem to the Turán number. In Sections 3 and 4 the main results of the paper and their proofs are presented, where Section 3 is an auxiliary section collecting general facts about the critical situation when for some n, k, v, $\mathcal{G}$ is not an optimal (n, k)-generator, but $\mathcal{G} - v$ is an optimal $(n-1, k)$-generator. Finally, in Section 5 we show that natural refinements of the problem in the spirit of combinatorial optimization are NP-hard, and prove on the other hand that the construction provides a generator that never exceeds a small constant times the optimum. In the Appendix we show some more results concerning the case $k = 2$, which enabled us to finish some more concrete particular cases of the conjecture.

A consequence of the results presented in this paper is as follows.

https://doi.org/10.1017/S096354830800922X Published online by Cambridge University Press

Theorem 1.1. *For $n \leqslant 3k$ and $(n,k) \in \{(7,2),(8,2)\}$,*

$$E(n,k) = k \times (2^{\lceil n/k \rceil} - 1) - (\lceil n/k \rceil\, k - n) \times 2^{\lceil n/k \rceil - 1}.$$

We conjecture that this formula is true for every value of n and k.

2. Construction

A natural way to construct a generator is to partition the set $[n]$ into k parts and to include all the non-empty subsets of each part in the generator. The cardinality of such a generator is minimum when the sizes of the parts differ by at most one.

More formally, let $p := p(n,k) := \lceil \frac{n}{k} \rceil$ and $r := r(n,k)$ such that $n = p\,k - r$ with $0 \leqslant r < k$. Let $V_1, \ldots, V_k$ be a partition of $[n]$ into r sets of size $p-1$ and $k-r$ sets of size p. The generator we are constructing for all $n,k \in \mathbb{N}$ is

$$\mathcal{Y}(n,k) := (\mathcal{P}(V_1) \cup \cdots \cup \mathcal{P}(V_k)) \backslash \{\emptyset\},$$

where V is an arbitrary set. The cardinality of such a generator is $Y(n,k) := r \times (2^{p-1} - 1) + (k-r) \times (2^p - 1)$. Note that

$$Y(n,k) = Y(n-1,k) + 2^{p-1},$$

and this simple recursive formula seems useful to keep in mind. It is sufficient to prove *the same recursive formula for* $E(n,k)$. For instance we have $Y(13,5) = 27$ for $n = 13$ and $k = 5$.

Clearly, $E(n,k) \leqslant Y(n,k)$, and in fact equality seems always to hold.

Conjecture 2.1. *For all $n,k \in \mathbb{N}$, the generator $\mathcal{Y}(n,k)$ is optimal.*

Quite surprisingly, this conjecture arose in production management, and for $k = 2$ it is a posthumous conjecture of Erdős.

Indeed, as Zoltán Füredi reports, Erdős [2], [3] asked the same question for $k = 2$ without requiring the disjointness of the sets. Could the same assertion be true for arbitrary k? Let $E^*(n,k)$ denote the optimum for this problem. Clearly, $E^*(n,k) \leqslant E(n,k) \leqslant Y(n,k)$, so if $E^*(n,k) = Y(n,k)$ is true for some (n,k), there is equality throughout for this (n,k). These equalities would mean that disjointness is an irrelevant requirement (in the sense that it does not change the optimum value). Could this be proved by some simple argument without necessarily settling the conjectures (see Conjecture 3.13)? In many results of the paper $E(n,k)$ can be replaced by $E^*(n,k)$: see the remarks at the end of Section 3.

Moreover, we also conjecture the unicity of the construction.

Conjecture 2.2. *For all $n,k \in \mathbb{N}$ such that $p(n,k) \neq 2$, $\mathcal{Y}(n,k)$ is the unique optimal (n,k)-generator.*

Trying to prove the preceding two conjectures inductively leads to the following conjecture that would imply both (see the next section).

For a hypergraph $\mathcal{G} \subseteq \mathcal{P}([n])$ and $z \in [n]$ let $\mathcal{G}(z) := \{g \in \mathcal{G} : z \in \mathcal{G}\}$.

https://doi.org/10.1017/S096354830800922X Published online by Cambridge University Press

Conjecture 2.3. *For all $n,k \in \mathbb{N}$, and for every (n,k)-generator $\mathcal{G}$, there exists $z \in [n]$ such that*

$$|\mathcal{G}(z)| \geqslant 2^{p(n,k)-1}.$$

We prove that Conjecture 2.3 is true for $p = 1, 2, 3$ and $(n,k) \in \{(7,2),(8,2)\}$ for which $p = 4$.

Notice that the partition underlying the construction is the same as that in Turán's theorem [6]. The two are actually related. The Turán number $T(n,s,l)$, where n, s, l are three positive integers with $l \leqslant s \leqslant n$, is the minimum number of subsets of size l of a set of size n, such that each subset of size s contains at least one of them. In a generator, since every subset of size $(l-1)k+1$ must contain a member of size at least l, there are at least $T(n,(l-1)k+1,l)$ members of size at least l.

Turán solved this problem for $l = 2$. If $l = 2$, that is $s = k+1$, his problem can be stated as follows: minimize the number of edges of a graph on n vertices so that the maximum number of pairwise non-adjacent vertices does not exceed k. Replacing every member g of a generator by a pair which is a subset of g, we always have this property. Turán proved that the unique minimum for this number is given by k cliques of almost equal size that partition the vertex-set. This partition coincides with the defining partition of the construction, showing that the number of members of size at least two in a generator is at least the number of sets of size exactly two in Turán's construction.

For $l \geqslant 3$, Turán conjectured that the partition into blocks still gives the solution to its problem, but this appears to be false. According to Sidorenko [5], for $n = 9$, $s = 5$, $l = 3$ with $k = 2$ and $s = (l-1)k+1$, Turán's construction provides $\binom{4}{3} + \binom{5}{3} = 14$ subsets of size 3 such that every 5-tuple contains at least one of them, whereas the affine plane of order 3 gives a solution with only 12 subsets with the same property. This example has been adapted by Füredi and Katona to find the minimum number of sets that 2-generate all 4-tuples of a set.

Indeed, for $n = 9$, the set of minimum size that 2-generates all 4-tuples can be defined with the help of the affine plane with $q = 3$: take the lines of two parallel classes (6 triplets) and the 2-element subsets of the lines for the two remaining parallel classes (9 pairs for each, in total 18). The generator $\mathcal{G}$ consisting of these 24 sets and the singletons 2-generate all the sets of size at most 4. Generalizing this construction, Füredi and Katona [3] prove that it provides the best estimate for 2-generating all 4-tuples for all n. Compare 24 with the size of the subset of $\mathcal{Y}(9,2)$ capable of achieving the same task, the 2- and 3-tuples of $\mathcal{Y}(9,2)$, $\binom{4}{3} + \binom{4}{2} + \binom{5}{3} + \binom{5}{2} = 30$. With 30 sets (adding to $\mathcal{G}$ the 6 lines of the affine space not yet included), the set of 5-tuples can also be generated.

We cannot continue in this direction, since finding the Turán number when $l \geqslant 3$ is known as a difficult open problem; moreover, a closer direct look using more than just the containments provides better lower bounds for the diversity problem in general (Section 5).

3. Induction

In this section we show some general facts that may help in inductive proofs provided we still have an optimal generator after the deletion of one or two elements. In order to analyse how $E(n,k)$ changes as a function of n, we need tight lower and upper estimates. The only upper

https://doi.org/10.1017/S096354830800922X Published online by Cambridge University Press

estimate we have is $Y(n,k)$ and we will use it all the time; in the lower estimates two parameters of a hypergraph will play a role, the degree and the minimum transversal and the like.

For a hypergraph $\mathcal{G} \subseteq \mathcal{P}([n])$ and a subset $Z \subseteq [n]$, we define

$$\begin{aligned}
\mathcal{G} - Z &:= \{g \in \mathcal{G} \ : g \cap Z = \emptyset\},\\
\mathcal{G}(Z) &:= \{g \in \mathcal{G} \ : Z \subseteq g\},\\
\mathcal{G} \sqcap Z &:= \{g \cap Z \ : g \in \mathcal{G}\},\\
\mathcal{G} \sqcup Z &:= \{g \cup Z \ : g \in \mathcal{G}\},\\
\mathcal{G}/Z &:= \{g \setminus Z \ : g \in \mathcal{G}\}.
\end{aligned}$$

One-element sets $Z = \{z\}$ are often replaced by z when the usage is evident. Let us see some examples of occurrences of $z \in [n]$ and $U \subseteq [n]$:

$$\begin{aligned}
\mathcal{G} - z &= \{g \in \mathcal{G} \ : z \notin g\} &&= \mathcal{G} \setminus \mathcal{G}(z),\\
\mathcal{G}/z &= \{g \setminus \{z\} \ : g \in \mathcal{G}\},\\
\mathcal{G}(z)/z &= \{g \setminus \{z\} \ : g \in \mathcal{G}(z)\},\\
\mathcal{G}(z) - U &= \{g \in \mathcal{G} \ : z \in g,\ g \cap U = \emptyset\},\\
\mathcal{G}(z) \sqcup U &= \{g \cup U \ : g \in \mathcal{G},\ z \in g\}.
\end{aligned}$$

The quantity $|\mathcal{G}(z)|$ is usually called the *degree* of z in the hypergraph $\mathcal{G}$. Note that $\mathcal{G}(z)/z = \mathcal{H}$ if and only if $\mathcal{G}(z) = \{z\} \sqcup \mathcal{H}$.

We will actually need to refine our sets and our quantities. For a hypergraph $\mathcal{G} \subseteq \mathcal{P}([n])$ and $p \in \mathbb{N}$, $i = 1, \dots p$, we let $\mathcal{G}^i := \{g \in \mathcal{G} \ : |g| \geqslant i\}$ and $Y^i(n,k) := |\mathcal{Y}^i(n,k)|$.

In $\mathcal{Y}(13,5)$ there are 13 hyperedges of size 1, 11 of size 2 and 3 of size 3, so $Y^1(13,5) = 27$, $Y^2(13,5) = 14$, $Y^3(13,5) = 3$; $Y^i(n,k) - Y^{i+1}(n,k)$ $(i = 1, \dots, p)$ is the number of members of size exactly i.

We should not hope for anything stronger than Conjecture 2.3, which already implies all the other conjectures. However, we may need more details for a proof (as will be the case for some of our results).

Conjecture 3.1. *For all $n,k \in \mathbb{N}$, and for every (n,k)-generator $\mathcal{G}$, we have*

$$|\mathcal{G}^i| \geqslant Y^i(n,k) \text{ for all } i = 1, \dots, p. \tag{3.1}$$

Since $Y^1(n,k) = Y(n,k)$ this conjecture contains Conjecture 2.1. When the average degree is not far from the maximum (if $n = pk$ or, more generally, when r is small compared with k) it also implies Conjecture 2.3.

Proposition 3.2. *If $n = pk$ and* (3.1) *holds for a hypergraph $\mathcal{G}$, then the average degree in $\mathcal{G}$ is at least 2^{p-1}, and every degree is equal to this number if and only if there is equality everywhere in* (3.1).

https://doi.org/10.1017/S096354830800922X Published online by Cambridge University Press

Proof. The average degree of $\mathcal{G}$ is equal to the sum of the sizes in $\mathcal{G}$ divided by n, which in turn is equal to

$$1/n \sum_{i=1}^{n} |\mathcal{G}^i|,$$

since a set of size s is encountered here for the values $i = 1, \ldots, s$, that is, exactly s times.

If (3.1) holds, then this number is greater than or equal to the average degree of the hypergraph $\mathcal{Y}(n,k)$, which is equal to 2^{p-1}, since all degrees are equal to this number. Therefore all degrees are equal to 2^{p-1} if and only if there is equality everywhere in (3.1), as claimed. □

Proposition 3.3. *For all $i = 1, \ldots, p$,*

$$\begin{aligned} |\{g \in \mathcal{Y}(n,k) : |g| = i\}| &= Y^i(n,k) - Y^{i+1}(n,k) \\ &= r\binom{p-1}{i} + (k-r)\binom{p}{i}. \end{aligned}$$

If $H \subseteq \mathcal{P}([n])$ is a hypergraph, a *transversal* is a set that meets all members of $\mathcal{H}$, and $\tau(\mathcal{H})$ denotes the minimum size of a transversal. If $\mathcal{H}$ has m disjoint members, then clearly $\tau(\mathcal{H}) \geqslant m$. If $\mathcal{H}$ contains the empty set, it has no transversal, and we then define $\tau(\mathcal{H}) = \infty$.

Generators can be characterized in term of transversals by the following easy but useful proposition.

Proposition 3.4. *Let $\mathcal{G} \subseteq \mathcal{P}([n])$ be an (n,k)-generator, and $i \in \{1, \ldots, p\}$. Then $\tau(\mathcal{G}^i) \geqslant k(p - i + 1) - r$, and this bound is tight.*

Proof. Suppose $\mathcal{G}$ is an (n,k)-generator, and $T \subseteq [n]$, $|T| < k(p-i+1) - r$. Then $|V - T| = n - |T| > kp - r - (k(p-i+1) - r) = k(i-1)$, so in a partition into k elements there is a part of size at least i, so T is not a transversal of $\mathcal{G}^i$, and the proposition is proved. Equality holds for $\mathcal{G} = \mathcal{Y}(n,k)$. □

The extreme case $i = p$ of Conjecture 3.1 is now easy and required.

Proposition 3.5. *If $\mathcal{G}$ is an (n,k)-generator, then $|\mathcal{G}^p| \geqslant Y^p(n,k) = k - r = n - (p-1)k$, and if equality holds, then $\mathcal{G}$ contains exactly $k - r$ sets of size at least p, and they are pairwise disjoint.*

Proof. Apply the preceding proposition to $i = p$: $|\mathcal{G}^p| \geqslant \tau(\mathcal{G}^p) \geqslant k - r = n - (p-1)k$, and if equality holds throughout, then in particular $|\mathcal{G}^p| = \tau(\mathcal{G}^p)$, that is, all the sets of $\mathcal{G}^p$ are pairwise disjoint. □

We now prove that *Conjecture 2.3 implies Conjecture 2.1 and Conjecture 2.2.*

The following lemma deduces the optimality of the construction – that is, Conjecture 2.1 – by induction on n if and only if there always exists an optimal generator containing a vertex of degree at least 2^{p-1} (which is somewhat weaker than Conjecture 2.3; see Conjecture 3.11 below).

https://doi.org/10.1017/S096354830800922X Published online by Cambridge University Press

Lemma 3.6. *Let $\mathcal{G}$ be an optimal (n,k)-generator, $z \in [n]$, $|\mathcal{G}(z)| \geqslant 2^{p-1}$, and assume $Y(n-1,k) = E(n-1,k)$. Then,*

$$|\mathcal{G}(z)| = 2^{p-1}, \quad Y(n,k) = E(n,k).$$

Proof. Since $\mathcal{G} - z$ generates $\mathcal{P}([n] \setminus \{z\})$, it is an $(n-1,k)$-generator:

$$\begin{aligned} E(n,k) = |\mathcal{G}| &= |\mathcal{G}(z)| + |\mathcal{G} - z| \geqslant 2^{p-1} + E(n-1,k) \\ &= 2^{p-1} + Y(n-1,k) = Y(n,k), \end{aligned}$$

so there is equality everywhere. □

As a consequence, we see that Conjecture 2.1 follows recursively for (n,k) if we know Conjecture 2.3 for all (n',k), $k \leqslant n' < n$.

This recursion raises the question of analysing 'the moment when a generator deviates from the construction, while n is increased and k is fixed'. (We will see that Conjecture 2.3 is true if $n \leqslant 3k$.) In the construction there are vertices z for which $\mathcal{Y}(n,k) - z$ is isomorphic to $\mathcal{Y}(n-1,k)$. The following theorem shows that $|\mathcal{G}(z)|$ with $\mathcal{G} - z = \mathcal{Y}(n-1,k)$ has to pay a 'high price' for essentially deviating from the construction.

If $\mathcal{H}$ is a hypergraph on $[n]$, $z \in [n]$ and $z \notin U \subseteq [n]$, we say that z *sees* U if $\mathcal{G}(z) \sqcap U = \mathcal{P}(U)$. Furthermore it *strongly sees* U if $\mathcal{G}(z) \supseteq \{z\} \sqcup \mathcal{P}(U)$.

Theorem 3.7. *Let $\mathcal{G} \subseteq \mathcal{P}([n])$ be an (n,k)-generator, $z \in [n]$, and suppose*

$$\mathcal{G} - z \subseteq \mathcal{P}(V_1) \cup \cdots \cup \mathcal{P}(V_k)$$

for a partition $\{V_1, \ldots, V_k\}$ ($V_i \neq \emptyset, i = 1, \ldots, k$) of $[n] \setminus \{z\}$. Then there exists $1 \leqslant i \leqslant k$ ($i = 1$, say) such that z sees V_1; moreover, if it does not strongly see V_1, then $|\mathcal{G}(z)| \geqslant 2^{|V_1|} + m - 1$, where $m := \min_{i=2,\ldots,k} |V_i|$.

Note that since $\mathcal{G} - z$ generates $[n] \setminus z$, in fact equality holds in the condition. Introduce the notation $\mathcal{U} := \{U \subseteq V_1, \{z\} \cup U \notin \mathcal{G}\}$. Then z does not strongly see V_1 if and only if $\mathcal{U} \neq \emptyset$; $\{z\} \in \mathcal{G}$ implies $\emptyset \notin \mathcal{U}$, and therefore $\mathcal{U}$ has a non-empty member which is inclusionwise minimal.

Proof. Suppose for a contradiction that the first part of the theorem is false, that is, for all $i \in \{1, \ldots, k\}$, there exists $\alpha_i \in \mathcal{P}(V_i) \setminus (\mathcal{G}(z) \sqcap V_i)$. Since $\{z\} \in \mathcal{G}$ we have $\emptyset \in \mathcal{G}(z) \sqcap V_i$, so $\alpha_i \neq \emptyset$ for all $i = 1, \ldots, k$.

Now let $Z := \{z\} \cup \alpha_1 \cup \cdots \cup \alpha_k$. The set Z is generated by at most k members of $\mathcal{G}$, exactly one of which (say g) contains z. Clearly, $g \cap V_i \subseteq \alpha_i$, and $g \neq \alpha_i$ because of the definition of α_i ($i = 1, \ldots, n$). So $Z \setminus g$ still contains an element from each V_i ($i = 1, \ldots, k$), and therefore cannot be generated by at most $k-1$ members of $\mathcal{G} - z \subseteq \mathcal{P}(V_1) \cup \cdots \cup \mathcal{P}(V_k)$. This contradiction proves the first part of the theorem. That is, we can now assume $G(z) \sqcap V_1 = \mathcal{P}(V_1)$, defining $\mathcal{U}$ as before the proof, and note that *if $g \in \mathcal{G}(z)$, $g \cap V_1 = U \in \mathcal{U}$ then g meets $[n] \setminus (V_1 \cup z)$.*

https://doi.org/10.1017/S096354830800922X Published online by Cambridge University Press

To prove the stronger inequality of the theorem, let $U \in \mathcal{U}$ be minimal in $\mathcal{U}$; as noted, $U \neq \emptyset$. Define

$$\mathcal{G}_{=U} := \{g \in \mathcal{G}(z) : g \cap V_1 = U\} = \mathcal{G}(z \cup U) - (V_1 \setminus U), \quad \text{and}$$
$$\mathcal{G}_{\subsetneq U} := \{g \in \mathcal{G}(z) : g \cap V_1 \subsetneq U\} = (G(z) - (V_1 \setminus U)) \setminus \mathcal{G}_{=U}.$$

Clearly, $\mathcal{G}_{=U} \cap \mathcal{G}_{\subsetneq U} = \emptyset$. Let $\tau := \tau(\mathcal{G}_{=U}/(U \cup z))$, that is, τ is the minimum size of a set disjoint of $U \cup z$ that meets each member of $\mathcal{G}_{=U}$. This minimum is finite since, as noted, each member of $\mathcal{G}_{=U}$ has an element outside U. Note also that $|\mathcal{H}| \geqslant \tau(\mathcal{H})$ holds whenever the latter is finite. Therefore we can suppose $\tau < m$ without loss of generality, since otherwise $|\mathcal{G}_{=U}| \geqslant \tau \geqslant m$, and

$$|\mathcal{G}(z)| = |\mathcal{G}(z) \setminus \mathcal{G}_{=U}| + |\mathcal{G}_{=U}| \geqslant (2^{|V_1|} - 1) + m, \tag{3.2}$$

and nothing else remains to be proved.

Claim. $|\mathcal{G}_{\subsetneq U}| \geqslant 2^{|U|} + 2^{m-\tau} - 2$.

Since $U \in \mathcal{U}$ is minimal, $z \sqcup (\mathcal{P}(U) \setminus U) \subseteq \mathcal{G}_{\subsetneq U}$, so we already know $2^{|U|} - 1$ elements of $\mathcal{G}_{\subsetneq U}$. It suffices to show now that $\mathcal{G}_{\subsetneq U}$ has at least $2^{m-\tau} - 1$ elements that meet $[n] \setminus V_1$.

Let C be a transversal of $\mathcal{G}_{=U}/(U \cup z)$, $|C| = \tau$. Then $C \subseteq V_2 \cup \cdots \cup V_k$. Now the condition of the theorem is satisfied for $\mathcal{G} - ((V_1 \setminus U) \cup C)$, with the same z, and with the partition $\{U, V_2 \setminus C, \ldots, V_k \setminus C\}$: we already know $U \neq \emptyset$, and because of $|C| = \tau < m$, $V_i \setminus C \neq \emptyset$, $(i = 2, \ldots, k)$.

Since $U \in \mathcal{U}$ and C is a transversal of $\mathcal{G}_{=U}/(U \cup z)$, $\mathcal{G}(z) - (V_1 \setminus U) \cup C = \mathcal{G}_{\subsetneq U} - C$. Since z does not see U, by the already proven first assertion of our theorem it does see $V_i \setminus C$ for some $i = 2, \ldots, k$. Let $i = 2$: $V_2 \setminus C$ has at least $m - \tau$ elements, and therefore $\mathcal{P}(V_2 \setminus C)$ has at least $2^{m-\tau} - 1$ non-empty members.

Using that z sees V_1, and then applying the claim and the inequality $2^{m-\tau} \geqslant m - \tau + 1$, we get

$$\begin{aligned} |\mathcal{G}(z)| &= |\mathcal{G}(z) \setminus (\mathcal{G}_{=U} \cup \mathcal{G}_{\subsetneq U})| + |\mathcal{G}_{=U}| + |\mathcal{G}_{\subsetneq U}| \\ &\geqslant 2^{|V_1|} - 2^{|U|} + \tau + 2^{|U|} + 2^{m-\tau} - 2 \\ &\geqslant 2^{|V_1|} + \tau + (m - \tau + 1) - 2 = 2^{|V_1|} + m - 1. \end{aligned}$$

□

The equality case of the bounds is worth analysing too, in the hope of improving the estimates: the gains allow us to deduce stronger bounds on the degree from weaker bounds, and thereby the optimality of $\mathcal{Y}(n,k)$ for some n and k. In the following proposition and corollary we will suppose $\mathcal{G} \subseteq \mathcal{P}([n])$ is an (n,k)-generator, $z \in [n]$, and $\mathcal{G} - z \subseteq \mathcal{P}(V_1) \cup \cdots \cup \mathcal{P}(V_k)$ for a partition $\{V_1, \ldots, V_k\}$ $(V_i \neq \emptyset, i = 1, \ldots, k)$ of $[n] \setminus \{z\}$; we let μ, m denote the smallest and the second-smallest size among the sizes $\{|V_i| : i = 1, \ldots, k\}$ of the partition classes.

Under the condition of the theorem a first estimate is $|\mathcal{G}(z)| \geqslant 2^{\mu}$, since z sees one of the classes. The theorem claims that there is equality in this bound if and only if z strongly sees one of the smallest classes.

It is interesting that the bound jumps from $2^{|V_1|}$ to $2^{|V_1|} + m - 1$ if z sees V_1 but does not strongly see it. What are the conditions of equality then?

https://doi.org/10.1017/S096354830800922X Published online by Cambridge University Press

Proposition 3.8. *Suppose* $\mathcal{G} \subseteq \mathcal{P}([n])$ *is an* (n,k)*-generator,* $z \in [n]$*, and* $\mathcal{G} - z \subseteq \mathcal{P}(V_1) \cup \cdots \cup \mathcal{P}(V_k)$ *for a partition* $\{V_1, \ldots, V_k\}$ $(V_i \neq \emptyset, i = 1, \ldots, k)$ *of* $[n] \setminus \{z\}$*. If* z *sees* V_1 *but does not strongly see it, that is,* $\mathcal{U} \neq \emptyset$*, then equality holds in the bound*

$$|\mathcal{G}(z)| \geqslant 2^{|V_1|} + m - 1, \tag{3.3}$$

if and only if there exists $1 \leqslant i \leqslant k$ $(i = 2,$ *say*$)$ *such that* $|V_2| = m$*,* $V_2 = \{v_1, \ldots, v_m\}$ *and, choosing the indices appropriately, one of the following statements is true.*

(i) *There exists* $U \subseteq V_1$ *such that, with* $\mathcal{U}_1 := (\mathcal{P}(U) \setminus \{U\})$ *and* $\mathcal{U}_2 = \{U \cup \{v_i\} : i = 1, \ldots, m\}$*, or* $\mathcal{U}_2 = \{U \cup \{v_i\} : i = 1, \ldots, m-1\} \cup \{\{v_m\}\}$*,*

$$\mathcal{G}(z)/z = \mathcal{U}_1 \cup \mathcal{U}_2.$$

(ii) $m = 2$*,* $\mathcal{U} \subseteq \mathcal{P}(V_1)$ *is arbitrary,* $g_{=U} := U \cup \{v_1\}$ $(U \in \mathcal{U})$*, and*

$$\mathcal{G}(z)/z = (\mathcal{P}(U) \setminus \mathcal{U}) \cup \{g_{=U} : U \in \mathcal{U}\} \cup \{v_2\}.$$

(iii) $m = 1$*,* $\mathcal{U}$ *is arbitrary, and* $\mathcal{G}(z)/z = (\mathcal{P}(U) \setminus \mathcal{U}) \cup \{g_{=U} : U \in \mathcal{U}\}$*, where* $g_{=U}$ *is the union of* U *and an arbitrary non-empty set of elements that form singleton classes.*

Proof. Suppose the condition of Theorem 3.7 is satisfied, and (3.3) is satisfied with equality. Then $m \leqslant \tau$, since $m > \tau$ would imply that (3.2) would also be satisfied with strict inequality, and then so would be the identical (3.3). To have equality in the claim, $\mathcal{G}(z)$ cannot contain a set that meets a partition-class of size bigger than m different from V_1.

Now consider $\mathcal{U}$ as in the proof, and let $U \in \mathcal{U}$. Let us now exploit the equalities in the inequalities of the proof of (3.3) in the proof of Theorem 3.7 from the end backwards: in order to have equality in (3.3), we need $2^{m-\tau} = m - \tau + 1$, and since $m - \tau \geqslant 0$, this holds if and only if $m - \tau = 1$, or $m - \tau = 0$. We will have to consider both the case $\tau = m$ and $\tau = m - 1$.

If $m > 2$ then $|\mathcal{G}_{=U'}| > 1$ for all $U' \in \mathcal{U}$, while in (3.2) we used the bound of 1 for all but one $U \in \mathcal{U}$. So strict inequality holds if $|\mathcal{U}| > 1$. If $|\mathcal{U}| = 1$ the equality can hold, and the two cases corresponding to $\tau = m$ and $\tau = m - 1$ are listed in (i).

If $m = 2$ and $\tau = m$, then again $|\mathcal{G}_{=U'}| > 1$ for all $U' \in \mathcal{U}$, and strict equality can hold only if $|\mathcal{U}| = 1$, already included in the previous case. However, if $m = 2$ and $\tau = m - 1$, then $|\mathcal{G}_{=U'}| = 1$ is possible for all $U' \in \mathcal{U}$, and precisely if the unique element of $\mathcal{G}_{=U'}$ is the $\mathcal{G}_{=U'}$ of (ii). So all the new cases where equality can occur for $m = 2$ are listed in (ii).

If $m = 1$, then as noticed, all sets in $\mathcal{G}(z)$ must be included in the union of V_1 and the partition classes of size m, that is, must be of the form given in (iii). It is easy to check that this is then sufficient: all sets of this form are (n,k)-generators. □

We get the following corollary from the theorem and the above analysis of the equality. Recall the notations p and m.

Corollary 3.9. $\mathcal{G} \subseteq \mathcal{P}([n])$ *is an* (n,k)*-generator,* $z \in [n]$*, and* $\mathcal{G} - z \subseteq \mathcal{P}(V_1) \cup \cdots \cup \mathcal{P}(V_k)$ *for some partition* $\{V_1, \ldots, V_k\}$ $(V_i \neq \emptyset, i = 1, \ldots, k)$ *of* $[n] \setminus \{z\}$*. Then*

$$|\mathcal{G}(z)| \geqslant 2^{p-1} + m \tag{3.4}$$

unless z *strongly sees one of the classes, or one of* (i)*,* (ii) *or* (iii) *holds.*

https://doi.org/10.1017/S096354830800922X Published online by Cambridge University Press

The following lemma states, in addition to the optimality of the construction, the unicity of optima – that is, Conjecture 2.2 – by induction on n if and only if every optimal generator contains a vertex of degree at least 2^{p-1} (which is still somewhat weaker than Conjecture 2.3; see Conjecture 3.12).

Lemma 3.10. *Let $\mathcal{G}$ be an optimal (n,k)-generator, $z \in [n]$, $|\mathcal{G}(z)| \geqslant 2^{p-1}$ and $p \geqslant 3$; assume that $\mathcal{Y}(n-1,k)$ is the unique optimal $(n-1,k)$-generator. Then $\mathcal{G} = \mathcal{Y}(n,k)$.*

Proof. By Lemma 3.6, $|\mathcal{G}(z)| = 2^{p-1}$, and $|\mathcal{G}| = Y(n,k)$, whence $\mathcal{G} - z = Y(n,k) - 2^{p-1} = Y(n-1,k)$, and then, by the condition, $\mathcal{G} - z = \mathcal{Y}(n-1,k)$.

So $\mathcal{G} - z = (\mathcal{P}(V_1) \cup \cdots \cup \mathcal{P}(V_k))\backslash\{\emptyset\}$, where $\{V_1, \ldots, V_k\}$ is a partition of $[n]$ into parts of size $p(n,k)$ and $p(n,k)-1$. By Theorem 3.7 one can choose V_1 so that either $\mathcal{G}(z)/z = \mathcal{P}(V_1)$, or $|G(z)| \geqslant 2^{|V_1|} + m - 1$ with $m = \min_{i=2,\ldots,k} |V_i| = p(n,k) - 1 \geqslant 2$.

In the first case, by optimality, V_1 is a class of size $p(n,k)-1$ such that $\mathcal{G} = \mathcal{Y}(n,k)$ follows. If, indirectly, the second case holds, then

$$2^{p-1} = |\mathcal{G}(z)| \geqslant 2^{p-1} + m - 1 \geqslant 2^{p-1} + 1,$$

and this contradiction finishes the proof. □

Modified as follows, Conjecture 2.3 becomes equivalent to Conjecture 2.1 by Lemma 3.6.

Conjecture 3.11. *For all $n,k \in \mathbb{N}$, there exists an optimal (n,k)-generator $\mathcal{G}$ and $z \in [n]$ such that*

$$|\mathcal{G}(z)| \geqslant 2^{p(n,k)-1}. \tag{3.5}$$

Modified as follows, Conjecture 2.3 becomes equivalent to Conjecture 2.2 by Lemma 3.10.

Conjecture 3.12. *For all $n,k \in \mathbb{N}$, for every optimal (n,k)-generator $\mathcal{G}$ there exists $z \in [n]$ such that* (3.5) *holds.*

We thus have the following implication between the conjectures:

$$\begin{aligned}
&\text{Conjecture 2.3} \implies \text{Conjecture 2.2} \implies \text{Conjecture 2.1},\\
&\text{Conjecture 3.1} \implies \text{Conjecture 2.1}\\
&\text{Conjecture 2.1} \iff \text{Conjecture 3.11},\\
&\text{Conjecture 2.2} \iff \text{Conjecture 3.12}.
\end{aligned}$$

Let us also state the conjecture asserting that the disjointness requirement does not change the optimum value.

Conjecture 3.13. *For all $n,k \in \mathbb{N}$, $E^*(n,k) = E(n,k)$.*

So far all the simple propositions, lemmas and conjectures hold without change if disjointness is not required and E^* is written instead of E. This is not true, though, for Theorem 3.7 and its

https://doi.org/10.1017/S096354830800922X Published online by Cambridge University Press

corollaries, including Lemma 3.10 and Proposition 3.8, the reason being that it was essential that at most one of the k disjoint sets contains a given $z \in [n]$.

4. Case $p \leqslant 3$

Recall the notation $p = p(n,k) = \lceil \frac{n}{k} \rceil$ and $n = pk - r$ with $0 \leqslant r < k$. In this section we prove all the conjectures for $p \leqslant 3$. This is done in Theorem 4.1 for $p \leqslant 2$, and in Theorem 4.3 for $p = 3$. (In the Appendix we add to this the two first cases with $p = 4$: $(n,k) = (7,2)$ and $(n,k) = (8,2)$.)

Theorem 4.1. *If $p \leqslant 2$, that is, $1 \leqslant n \leqslant 2k$, then $E^*(n,k) = E(n,k) = Y(n,k)$, furthermore, for any (not necessarily optimal) (n,k)-generator $\mathcal{G}$,* (3.1) *holds, and there exists $z \in [n]$ such that* (3.5) *holds. A generator $\mathcal{G}$ is optimal if and only if it consists of all the singletons in $[n]$ and $n-k$ pairwise disjoint sets of size at least 2.*

In particular, the construction is the unique optimal generator if $n \leqslant k$ or $n = 2k$, but it is not unique if $k < n < 2k$. However, if $k < n < 2k$, Conjecture 3.1 still follows easily, and it is also not an exception of Theorem 3.7 or the reformulation of its essential part in Lemma 3.10, useful for proving unicity; this case is an exception to unicity only because, for $m = 1$ (and only in this case), Theorem 3.7 does not exclude other optimal solutions of the same size, and they do indeed exist and are already mentioned in case (iii) of equality. The reason for this is simply that the validity of $2^{p-1} + m - 1 = 2^{p-1}$ in this case.

This is also the only case when 'Turán's bound' $T(n,k+1,2)$ is exact.

Proof. Let $\mathcal{G}$ be an arbitrary (n,k)-generator. It contains all the singletons, and if $p = 1$, that is, $n \leqslant k$, there is no need of more members.

If $p = 2$, that is, $k+1 \leqslant n \leqslant 2k$, then by Proposition 3.5, $|\mathcal{G}^2| \geqslant Y^2(n,k) = k - r = n - k$, and equality holds if and only if the sets of size at least 2 are disjoint.

Conversely, suppose the hypergraph $\mathcal{G}$ has $n-k$ disjoint members of size at least 2 ($k+1 \leqslant n \leqslant 2k$), and let us check that it is an (n,k)-generator. Let $S \subseteq [n], s := |S| > k$. Then S misses at most $n - s < k$ members of $\mathcal{G}^2$, so it contains at least $n - k - (n-s) = s - k$ members of $\mathcal{G}^2$, all pairwise disjoint. So S can be generated by $s-k$ members of $\mathcal{G}^2$, plus at most $s - 2(s-k) = 2k - s$ singletons. □

Theorem 4.2. *If $p = 3$, that is, $2k < n \leqslant 3k$, then for any (not necessarily optimal) (n,k)-generator $\mathcal{G}$,* (3.1) *holds.*

Proof. We have (3.1) for $i = 3$ by Proposition 3.5: $|\mathcal{G}^3| \geqslant Y^3(n,k) = n - 2k$.

Now we prove (3.1) for $i = 2$, by induction on $n - 2k$. By Theorem 4.1 it is true for $n = 2k$. For the sake of easier understanding, we first do the proof separately for $n = 2k+1$, using it for $n = 2k$. For all $z \in [2k+1]$ we have $|\mathcal{G}^2 - z| \geqslant Y^2(2k,k) + 1 = k + 1$; otherwise we are done by Lemma 3.10. Now

$$\sum_{z \in [n]} |\mathcal{G}^2 - z| \geqslant (2k+1)(k+1),$$

https://doi.org/10.1017/S096354830800922X Published online by Cambridge University Press

and in this sum every member of G^2 is counted at most $2k-1$ times, so

$$|\mathcal{G}^2| \geqslant \frac{2k+1}{2k-1}(k+1) = \frac{k+1/2}{k-1/2}(k+1) > k+2.$$

(For an easier look at it we have used the fact that, multiplying a number x by $\frac{k+1/2}{k-1/2}$, it increases by more than 1 if and only if $x > k-1/2$.) Since $Y^2(2k+1,k) = Y^2(2k,k)+3$, $|\mathcal{G}^2| \geqslant k+3 = Y^2(2k,k)+3 = Y^2(2k+1,k)$, as claimed.

Similarly, for an arbitrary (n,k)-generator, $2k+1 \leqslant n \leqslant 3k$, we have

$$|\mathcal{G}^2| \geqslant \frac{n}{n-2}\big(E(n-1,k)-(n-1)+1\big) > E(n-1,k)-(n-1)+2,$$

since $E(n-1,k) > \frac{n-2}{2}$, and the statement then follows using $Y^2(n,k) = Y^2(n-1,k)+3$. □

We do not see how to deduce Conjecture 2.3 from the above theorem. On the other hand, we can prove this conjecture separately (for $p=3$), implying the previous theorem as well, in a simpler way, and without using any of the previous results or the disjointness of generators. For $i=3$, (3.5) is easy, and the following theorem implies it for $i=2$ and $i=1$. For $2k \leqslant n \leqslant 3k$ we will thus have two proofs of the optimality. (We included the previous theorem because it forecasts our future difficulties: whenever the average degree of $\mathcal{Y}(n,k)$ is much smaller than the maximum degree, 'averaging arguments' do not easily work.)

Theorem 4.3. *If $p=3$, that is $2k < n \leqslant 3k$, then $E^*(n,k) = E(n,k) = Y(n,k)$; furthermore, for any (not necessarily optimal) (n,k)-generator $\mathcal{G}$, (3.1) holds, and there exists $z \in [n]$ such that (3.5) holds. The construction is the unique (n,k)-generator.*

Proof. We prove, without requiring disjointness, that for any (n,k)-generator $\mathcal{G}$ there exists $z \in [n]$ such that (3.5) holds.

We can suppose without loss of generality that $n = 2k+1$. Indeed, if $n > 2k+1$, then we can apply the proven assertion to the $(2k+1,k)$-generator $\mathcal{G}(U)$, where $U \subseteq [n]$, $|U| = 2k+1$.

Let $\mathcal{G}$ be an (n,k)-generator, and suppose for a contradiction that $|\mathcal{G}^2(z)| \leqslant 2$ for all $z \in [n]$.

We define an undirected graph G on vertices $V := [n] = [2k+1]$, in the following way: for each $g \in \mathcal{G}^2$, we choose two distinct vertices $u,v \in g$, let $e = uv$ be an edge of G, and use the notation g_e for g. For $g_1 \neq g_2 \in \mathcal{G}^2$ we can take the same u,v (if $u,v \in g_1 \cap g_2$), but then we take two parallel uv edges e_1 and e_2. We will say that the edge $e = uv$ *represents* $g_e \in \mathcal{G}^2$. We thus suppose that different sets in $\mathcal{G}^2$ are represented by different edges. Furthermore, we suppose that we make the possible choices of u and v so as to minimize the number of components of G.

Now it follows from the indirect assumption that all the degrees of the graph G are at most 2, so G is a disjoint union of cycles, paths and isolated vertices. The following claim is the key to the proof.

Claim. Let C be a cycle of G, and let e be an edge of C. Then $e \in \mathcal{G}^2$, and is not contained in any bigger set of $\mathcal{G}^2$.

Indeed, by the definition of G, e is contained in a set of $\mathcal{G}^2$, so it is sufficient to prove that no set in $\mathcal{G}^2$ can properly contain e. If there exists such a set, then it is g_e; otherwise the endpoints of e would be contained in three different sets of $\mathcal{G}^2$.

https://doi.org/10.1017/S096354830800922X Published online by Cambridge University Press

- If an extra element z of g_e (different from the endpoints of e) of such a set were in C, then z would be contained in three different sets of $\mathcal{G}^2$: g_a and g_b, where a, b are the two edges incident to z in C, and $g_e \supseteq e \cup \{z\}$. Clearly, e, a, b are different, and therefore g_e, g_a, g_b as well, contradicting the indirect assumption.
- If an extra element z of g_e (different from the endpoints of e) of such a set were in another component K of G, then replacing one of the endpoints of e by a point in $g_e \cap K$, we get another representation of $\mathcal{G}$ with one fewer component (all vertices of C and K are now in the same component), contradicting the definition of G.

The claim is proved.

Let U be the set of vertices of G that are in a cycle. The subgraph $G(V \setminus U)$ contains only paths and isolated vertices, so we can find a stable set (not containing both endpoints of an edge) S of $G(V \setminus U)$ such that $|S| \geqslant |V \setminus U|/2$. (We take a (the) bigger stable set in each component.)

We show now that $S \cup U$ cannot be k-generated, contradicting the choice of $\mathcal{G}$. Recall that any $g \in \mathcal{G}^2$, $g \subseteq S$ has also an edge in G. But the only edges in $S \cup U$ are in the cycles, and for these the claim holds. Therefore what we have to show is exactly that $S \cup U$ is not the union of at most k edges of G or singletons.

Indeed, let $\gamma(X)$ denote the minimum number of edges and singletons necessary for generating a set $X \subseteq n$. Let the components of G be $C_1, \ldots, C_t$ ($t \in \mathbb{N}$). Note that for all $i = 1, \ldots, t$, $\gamma(U \cap C_i) \geqslant |C_i|/2$. Then

$$\gamma(U) = \sum_{i=1}^{t} \gamma(U \cap C_i) \geqslant \sum_{i=1}^{t} |C_i|/2 = \frac{2k+1}{2} > k.$$

So U cannot be k-generated, a contradiction.

By Lemma 3.6 (which does not require disjointness), $E^*(n,k) = E(n,k) = Y(n,k)$ follows. When disjointness is required, by Lemma 3.10, the construction is the unique optimal (n,k)-generator. □

5. Optimization and approximation

The general problem this work is concerned with is a natural question in combinatorial optimization, also including computational complexity and approximation ratios. In this section we would like to present our related observations: some negative results concerning the computational complexity, and simple but surprisingly good estimates for the quantity $E(n,k)$.

Two natural optimization problems arise.

- We do not want to generate all cars, that is, all subsets of options, just a pre-given family.
- The generator is restricted to choosing elements from a given hypergraph.

More precisely, we have the following.

Problem: CHOOSY CUSTOMER'S DIVERSITY.
Input: $\mathcal{C} \subseteq \mathcal{P}([n])$, numbers k, s.
Question: Does there exist $\mathcal{G} \subseteq \mathcal{P}([n])$ that k-generates all sets in $\mathcal{C}$, and $|\mathcal{G}| \leqslant s$?

https://doi.org/10.1017/S096354830800922X Published online by Cambridge University Press

Problem: CONSTRAINED PRODUCER'S DIVERSITY.
Input: $\mathcal{H} \subseteq \mathcal{P}([n])$, number k and a target-set $T \subseteq [n]$.
Question: Does there exist $\mathcal{G} \subseteq \mathcal{H}$ that k-generates T?

Note that in this second problem we only speak of the *existence* of a generator. These are just two simple and natural variants that we choose for the sake of examples. The reader may enjoy stating his favourite variants and checking NP-completeness for them.

Theorem 5.1. *Both the* CHOOSY CUSTOMER'S *and* CONSTRAINED PRODUCER'S DIVERSITY *problems are NP-complete.*

Proof. We first reduce VERTEX COVER to CHOOSY CUSTOMER'S DIVERSITY, and even to instances where $k = 2$. (VERTEX COVER and 3DM below are proved to be NP-complete in Garey and Johnson's seminal book [4].)

Let $G = (V, E)$ be a graph, and consider the problem with input $\Omega = V \cup \{u\}$, where u is an extra vertex not in V, and $\mathcal{C} := \{\{v\} : v \in \Omega\} \cup \{\{a, b, u\} : a, b \in V, ab \in E\}$.

Clearly, if T is a vertex cover, that is, $T \cap e \neq \emptyset$ for all $e \in E$, then $\mathcal{G} := \{\{v\} : v \in \Omega\} \cup \{\{t, u\} : t \in T\}$ does 2-generate all $C \in \mathcal{C}$. Conversely, $\{\{v\} : v \in \Omega\}$ must be contained in all generators, and all the other sets can be supposed to contain u and to be of size 2. (Otherwise we can add u and keep only one of the elements different from u.) Let $T := \{v \in V : (v, u) \in \mathcal{G}\}$. Then T is a vertex cover, finishing the proof of the first assertion.

Let us now reduce 3DM to CONSTRAINED PRODUCER'S DIVERSITY. Let (U, V, W, E) be an instance of 3DM, that is, $E \subseteq U \times V \times W$ (the Cartesian product of U, V, W), where $|U| = |V| = |W| = 3k$. Define $T := U \cup V \cup W$. Now, clearly, $\mathcal{G} \subseteq E$ k-generates T if and only if it is a 3-dimensional matching (that is, if and only if it partitions T). □

In both proofs it is irrelevant whether or not we ask disjointness from the generators. (In these cases there exists always a disjoint optimal solution.)

We now show that the construction provides a quite good approximation of the optimum. Enumeration provides the bound $Y(n,k) \leqslant E(n + 2k, k)$. Let us sketch a proof of this. Given an (n,k)-generator $\mathcal{G}$, all the $2^n - 1$ non-empty subsets of $[n]$ can be encoded by an at most k-element subset of $\mathcal{G}$:

$$\sum_{i=1}^{k} \binom{|\mathcal{G}|}{i} \geqslant 2^n - 1.$$

It follows that $k|\mathcal{G}|^k/k! \geqslant 2^n$, that is, $|\mathcal{G}|^k \geqslant (k-1)!\,2^n$, and applying Stirling's formula and taking the kth root: $|\mathcal{G}| \geqslant \frac{k-1}{e} 2^{n/k}$. So $E(n,k) \geqslant \frac{k-1}{e} 2^{n/k}$, while $Y(n,k) \leqslant k2^{n/k} + \mathcal{O}(1)$, which shows that $Y(n,k)/E(n,k)$ does not exceed $\varepsilon(n,k)e$, where $\lim_{n,k\to\infty} \varepsilon(n,k) = 1$. The exact threshold valid for all n and k is certainly smaller than 4: $Y(n,k) \leqslant 4\,E(n,k)$. Since $Y(n + 2k, k) \geqslant 4Y(n,k)$, we get that $Y(n,k) \leqslant E(n + 2k, k)$.

For small k we do not have to apply Stirling's formula and we get essentially better bounds: for $k = 2$, we get $|\mathcal{G}| + \binom{|\mathcal{G}|}{2} \geqslant 2^n - 1$ and we get the same bounds as in the theorems below. Still with the same method, for $k = 3$ we get that the construction is at most

https://doi.org/10.1017/S096354830800922X Published online by Cambridge University Press

$\frac{4}{\sqrt[3]{12}} = 1,747\cdots$ times the optimum. Let us deduce the results for $k = 2$ with another method as well, which will also lead to a simple general proposition for arbitrary k.

Theorem 5.2. *For all $n \in \mathbb{N}$,*

$$E(n,2) \leqslant Y(n,2) \leqslant 3/2E(n,2),$$

and the constant 3/2 can actually be improved to $\sqrt{2}$ if n is even.

Expressing $E(n,2)$, $CY(n,2) \leqslant E(n,2) \leqslant Y(n,2)$, with $C = 2/3$ if n is odd, and $C = \sqrt{2}/2$ if n is even.

Proof. Let $\mathcal{G}$ be an $(n,2)$-generator. Since every subset of $[n]$ containing z is the union of a set in $\mathcal{G}(z)$ and a set in $(\mathcal{G} - z) \cup \{\emptyset\}$, we have

$$|\mathcal{G}(z)|(|\mathcal{G} - z| + 1) \geqslant 2^{n-1}.$$

The minimum of $x + y$, $(x, y \in \mathbb{R})$ under the condition $xy = 2^{n-1}$ is $x = y = 2^{\frac{n-1}{2}}$. Therefore, if in addition $\mathcal{G}$ is an optimal (n,k)-generator, then

$$E(n,2) = |\mathcal{G}| = |\mathcal{G}(z)| + |\mathcal{G} - z| \geqslant \min\{x + y - 1 : xy = 2^{n-1}\} = 2^{\frac{n-1}{2}} + 2^{\frac{n-1}{2}} - 1.$$

On the other hand, $Y(n,2) = 2^{\frac{n-1}{2}} - 1 + 2^{\frac{n+1}{2}} - 1$ if n is odd, and $Y(n,2) = 2^{\frac{n}{2}} - 1 + 2^{\frac{n}{2}} - 1$ if n is even. □

If we compare $Y(n,2)$ with the same estimates applied to $E(n+1,k)$ or $E(n+2,k)$, we get the following.

Theorem 5.3. *For all $n \in \mathbb{N}$, $E(n,2) \leqslant Y(n,2) \leqslant E(n+1,2)$, if n is even, and $E(n,2) \leqslant Y(n,2) \leqslant 3/4E(n+2,2)$, if n is odd.*

Finally, we now use the same method to prove a general statement previously of only isolated interest. For a hypergraph $\mathcal{H}$ let

$$\alpha(\mathcal{H}) := \max\{|S| : S \subseteq [\mathrm{n}], H \cap S \text{ is a singleton for all } H \in \mathcal{H}\}.$$

Note that an (n,k)-generator $\mathcal{G}$ always satisfies $\alpha(\mathcal{G}) \leqslant k$. On the other hand, for all n, k, $\alpha(\mathcal{Y}(n,k)) = k$. Conversely, a generator $\mathcal{G}$ with $\alpha(\mathcal{G}) = k$ looks close to the optimum, and we can easily prove that it is optimal if $n = pk$.

Proposition 5.4. *Suppose $n = pk$, $E(n-k,k) = Y(n-k,k)$, and that there exists an optimal (n,k)-generator $\mathcal{G}$, $\alpha(\mathcal{G}) = k$. Then $E(n,k) = Y(n,k)$ and $\mathcal{Y}(n,k)$ is the unique optimal (n,k)-generator, provided the same holds for $(n-k,k)$.*

If $k = 2$, the condition is $\alpha(\mathcal{G}) = 2$ and this means $x, y \in [n]$ such that $\mathcal{G}(x)$ and $\mathcal{G}(y)$ have no common elements. The proposition confirms all the conjectures under this condition (which is true for $\mathcal{Y}(n,2)$).

https://doi.org/10.1017/S096354830800922X Published online by Cambridge University Press

Let S be a set that meets all members of $\mathcal{H}$ only in one element, $|S| = k$. We will actually show that at least $Y(n,k)$ sets are needed *only to generate all sets in* $S \sqcup \mathcal{P}([n] \setminus S)$ *and in* $\mathcal{P}([n] \setminus S)$.

Proof. Clearly, any set containing S is generated by exactly k sets, exactly one from each $\mathcal{G}(s)$ ($s \in S$). Thus

$$\prod_{s\in S} |\mathcal{G}(s)| \geqslant 2^{n-k}.$$

By the inequality between the geometric and arithmetic means, we have under this condition

$$\sum_{s\in S} |\mathcal{G}(s)| \geqslant k2^{\frac{n-k}{k}} = k2^{p-1}. \tag{5.1}$$

Equality holds in (5.1) if and only if $|\mathcal{G}(s)| = 2^{\frac{n-k}{k}}$, and the members of $\cup_{s\in S}\mathcal{G}(s)$ generate $\prod_{s\in S} |\mathcal{G}(s)|$ sets; the latter condition holds if and only if any pair of sets from different $\mathcal{G}(s)$ are disjoint.

Define for all $s \in S$, $P_s := \cup\mathcal{G}(s)$. Because of $|\mathcal{G}(s)| = 2^{\frac{n-k}{k}}$ we have $|P_s \setminus \{s\}| \geqslant \frac{n-k}{k}$, that is,

$$\sum_{s\in S} |P_s| \geqslant k\left(\frac{n-k}{k} + 1\right) \geqslant n,$$

and if there is equality in (5.1) and therefore the sets P_s are pairwise disjoint, then there is equality everywhere, that is, $|P_s| = \frac{n-k}{k} + 1 = n/k = p$ for all $s \in S$.

We have now reached our final estimation, one ingredient of which is (5.1), and the other is the obvious inequality $|\mathcal{G} - S| \geqslant E(n-k,k)$. Then

$$\begin{aligned} E(n,k) = |\mathcal{G}| &= |\mathcal{G} - S| + \sum_{s\in S} |\mathcal{G}(s)| \\ &\geqslant E(n-k,k) + k2^{p-1} \\ &= Y(n-k,k) + k2^{p-1} = Y(n,k). \end{aligned}$$

So $E(n,k) = Y(n,k)$, and there is equality everywhere, so $\mathcal{G} - S$ is optimal. If $\mathcal{Y}(n-k,k)$ is the unique optimal $(n-k,k)$-generator then $\mathcal{G} - S$ is isomorphic to $\mathcal{Y}(n-k,k)$. Finally, applying Lemma 3.10 k times one by one to the elements of s in the role of z, we see that $\mathcal{G} = \mathcal{Y}(n,k)$. □

Conclusion. We proved that the most natural construction for an (n,k)-generator is optimal if $n \leqslant 3k$, and for some other individual pairs (n,k), regardless whether the disjointness of the sets is required; moreover, it is always a constant time approximation with a small constant. The natural formulations as an optimization problem are NP-hard.

Appendix: $k = 2$ and can we go further?

We deduce the conjecture for two more cases, also in order to provide another example of applying the arguments and assertions of the paper, and to realize the limits of some arguments.

The following lemma extends the validity of Theorem 3.7 to the case when $\mathcal{G} - z$ can contain one more set besides subsets of the partition classes. We restrict ourselves to the case $k = 2$ (the statement and its use seem to be considerably more complicated, even if not hopeless, for $k > 2$).

https://doi.org/10.1017/S096354830800922X Published online by Cambridge University Press

Lemma A1. *Suppose* $\mathcal{G}$ *is an* $(n,2)$*-generator,* $(\mathcal{G}-z)\subseteq \mathcal{P}(V_1)\cup\mathcal{P}(V_2)\cup\{h\}$*, where* $\{\{z\},V_1,V_2\}$ $(z\in[n])$ *is a partition of* $[n]$*,* $2\leqslant\mu:=|V_1|\leqslant|V_2|$*,* $h\subseteq[n]$*. Then* $|\mathcal{G}(z)|\geqslant 2^{\mu}$*; in particular,* $\mathcal{G}$ *is not optimal.*

Of course, we can suppose without loss of generality $h\cap V_i\neq\emptyset$ $(i=1,2)$, otherwise h can be omitted from $\mathcal{G}$, and the assertion follows from Theorem 3.7.

Proof. If z sees V_1 or V_2 we are done, so we suppose it does not.

Claim. For both $i=1$ and $i=2$, there is at most one subset of V_i that is not in $\mathcal{G}(z)\sqcap V_i$.

Suppose for a contradiction that the statement does not hold, say, for $i=2$: let $B\neq C\subseteq V_2$, $B,C\notin\mathcal{G}(z)\sqcap V_2$. Since z does not see V_1, there exists $A\subseteq V_1$, $A\notin\mathcal{G}(z)\sqcap V_1$. We show then $|\mathcal{G}(z)|\geqslant 2^{|V_1|}$.

The sets $\{z\}\cup A\cup B$, $\{z\}\cup A\cup C$ must contain h that must participate in 2-generating these sets, whence

$$\{z\}\cup(A\cup B)\setminus\{h\},\{z\}\cup(A\cup C)\setminus\{h\}\in\mathcal{G}.$$

We show now that $|\mathcal{G}(z)|\geqslant 2^{|V_1|}$, by labelling each subset of V_1 with a different set in $\mathcal{G}(z)$.

If $U\subseteq V_1$, $U\in\mathcal{G}(z)\sqcap V_1$, we label U with an arbitrary $g\in\mathcal{G}(z)$, $g\cap V_1=U$. For instance we label $\emptyset$ with $\{z\}$. If $A\notin\mathcal{G}(z)\sqcap V_1$, we saw that there exist two sets, $\{z\}\cup(A\cup B)\setminus\{h\},\{z\}\cup(A\cup C)\setminus\{h\}\in\mathcal{G}$. At most one of them is the label of $A\setminus\{h\}$; the other, say $(A\cup C)\setminus\{h\}$, is *a priori* not a label, since it meets V_1 also in $A\setminus\{h\}$, but it is not the label of this set. Let the label of A be $(A\cup C)\setminus\{h\}$. Clearly, the label of a different set $A'\subseteq V_1$, $A'\notin\mathcal{G}(z)\sqcap V_1$ is different, since it is $A'\cup C\setminus\{h\}$, different from $A\cup C\setminus\{h\}$. (Both $A\cup C$ and $A'\cup C$ contain h.) The claim is proved.

The claim implies that $|\mathcal{G}(z)|\geqslant 2^{|V_1|}-1$, but we are still fighting for strict inequality here. Let $1\in h\cap V_1$, $2\in h\cap V_2$. By Theorem 3.7, z sees $V_1\setminus\{1\}$ and $V_2\setminus\{2\}$ (since it does not see V_2 and V_1). If it strongly sees both of them, then $z\sqcup(V_1\setminus\{1\}),z\sqcup(V_2\setminus\{2\})\subseteq\mathcal{G}$, and the only common element of these two is z, so the bound is largely satisfied. If not, then in Theorem 3.7 equality is not satisfied, so there exists $A\subseteq V_1$ and $f,g\in\mathcal{G}$ such that $A=f\cap V_1=g\cap V_1$ for $f\neq g\in\mathcal{G}$, so the equality $|\mathcal{G}(z)|=2^{|V_1|}-1$ does not hold. □

Theorem A2. *For any* (*not necessarily optimal*) $(7,2)$*-generator,* (3.1) *holds, and* $\mathcal{Y}(7,2)$ *is the unique optimal generator.*

Proof. We first prove the second assertion. Let $\mathcal{G}$ be an optimal $(7,2)$-generator. Then $|\mathcal{G}|\leqslant Y(7,2)$. Add to $\mathcal{G}$ some new sets to get a hypergraph $\hat{\mathcal{G}}$ with $|\hat{\mathcal{G}}|=Y(7,2)=22$. Obviously $\hat{\mathcal{G}}$ is still a generator. It suffices to prove now that $\hat{\mathcal{G}}=\mathcal{Y}(7,2)$. Indeed, then $\mathcal{Y}(7,2)=\hat{\mathcal{G}}=\mathcal{G}$ follows since $\hat{\mathcal{G}}$ does not contain any other generator properly.

Let $d:=1/n\sum_{x\in[n]}\hat{\mathcal{G}}(x)$ be the *average degree* of $\mathcal{G}$. Clearly (as before, see Proposition 3.2),

$$dn=\sum_{x\in[n]}|\hat{\mathcal{G}}(x)|=\sum_{g\in\hat{\mathcal{G}}}|g|=\sum_{i=1}^{n}\hat{\mathcal{G}}^i. \tag{A1}$$

Claim 1. $d>6$

https://doi.org/10.1017/S096354830800922X Published online by Cambridge University Press

We already know $|\hat{\mathcal{G}}^1| \geqslant 22$ and therefore $|\hat{\mathcal{G}}^2| \geqslant 15$ as well. At the other extreme $|\hat{\mathcal{G}}^4| \geqslant 1$ is obvious; let $A \in \hat{\mathcal{G}}^4$. We show $|\hat{\mathcal{G}}^3| \geqslant 5$.

- If there exists $z \notin A$, $|\hat{\mathcal{G}}^3(z)| \geqslant 2$, then apply Proposition 3.4 after deleting z: $|\hat{\mathcal{G}}^3 - z| \geqslant \tau(\hat{\mathcal{G}}^3 - z) \geqslant 2$. But this bound is self-improving: $|A| \geqslant 4$, so A is not disjoint of the other set in $\hat{\mathcal{G}}^3 - z$, and therefore $|\hat{\mathcal{G}}^3 - z| \geqslant \tau(\hat{\mathcal{G}}^3 - z) + 1 \geqslant 3$. But then $|\hat{\mathcal{G}}^3(z)| + |\hat{\mathcal{G}}^3 - z| \geqslant 2 + 3 = 5$.
- If there exists $z \in A$, $|\hat{\mathcal{G}}^3(z)| \geqslant 3$, then similarly, apply simply $|\hat{\mathcal{G}}^3 - z| \geqslant \tau(\hat{\mathcal{G}}^3 - z) \geqslant 2$ to get $|\hat{\mathcal{G}}^3(z)| + |\hat{\mathcal{G}}^3 - z| \geqslant 3 + 2 = 5$.
- One of the preceding cases holds, because otherwise every $z \in [n]$ is covered by at most one member of $\hat{\mathcal{G}}^3 \setminus \{A\}$, although there are at least 3 sets of size at least 3 in this hypergraph on 7 elements.

We now conclude the proof of the claim by equation (A1):

$$d \geqslant \frac{22 + 15 + 5 + 1}{7} = \frac{43}{7} > 6.$$

According to the claim there exists $x \in [n]$, $|\hat{\mathcal{G}}(x)| \geqslant 7$, that is, $|\hat{\mathcal{G}} - x| \leqslant 22 - 7 = 15 = Y(6,2) + 1$. If strict inequality holds, we are done by Lemma 3.10, so we can suppose $|\hat{\mathcal{G}}(x)| = 7$.

Now Proposition 3.2 can be applied for $n = 6, k = 2, p = 3$: there exists $z \in [n]$, $\hat{\mathcal{G}}(z) - x \geqslant 2^{p-1} + 1$. So $|\hat{\mathcal{G}} - \{x, z\}| \leqslant Y(5,2)$, and equality holds here by Theorem 4.1. Now Theorem 3.7 can be applied to deduce that z strongly sees the class of size 2 of $\hat{\mathcal{G}} - \{x, z\}$, since $m = 3$. So $\hat{\mathcal{G}} - x$ contains a hypergraph isomorphic to $\mathcal{Y}(6,2)$, meaning that it is exactly $\mathcal{Y}(6,2)$ and one more element h. We now conclude the second part of the theorem with Lemma A1 substituting z for x.

Now let $\mathcal{G}$ be an arbitrary $(7,2)$-generator. By the already proven part we have (3.1) for $i = 1$ and $i = 2$. It is also obvious for $i = 4$; as above, denote $A \in \mathcal{G}^4$. In exactly the same way as we proceeded above, we can get $|\mathcal{G}^3| \geqslant 5$, after which it is still possible to do one more self-improving step, to prove $|\mathcal{G}^3| \geqslant 6 = Y^3(7,3)$, as claimed.

Suppose for a contradiction $|\mathcal{G}^3| \leqslant 5$. A set $T \in \mathcal{G}$, $|T| = 3$ will be called a *triangle*.

Claim 2. If $|\mathcal{G}^3(z)| \geqslant 3$ then $\mathcal{G}^3 - z$ has exactly two disjoint triangles, and these partition $[n] \setminus z$.

Indeed, $|\mathcal{G}^3 - z| \geqslant \tau(\mathcal{G}^3 - z) \geqslant 2$, and if one of these two inequalities is strict, then we arrive at the contradiction $5 \geqslant |\mathcal{G}^3| = |\mathcal{G}^3(z)| + |\mathcal{G}^3 - z| \geqslant 3 + 3 = 6$.

The average degree of $\mathcal{G}^3$ is at least $\frac{4+3+3+3+3}{7} = 16/7 > 2$. So there exists $z \in \mathcal{G}$, $|\mathcal{G}^3(z)| \geqslant 3$, and Claim 2 can be applied. Let T_1 and T_2 be the two triangles of $G^3 - z$ provided by Claim 2. Since A is not a triangle, it does not coincide with any of these, so $z \in A$. Let $T_3 \neq T_4 \in \mathcal{G}(z) \setminus \{A\}$.

Claim 3. $T_3 \cap T_4 = \{z\}$.

Indeed, another common element of T_3 and T_4 (denote it by x) would also be contained in T_1 or T_2, say T_1. Then $x \in T_1 \cap T_3 \cap T_4$, and also $x \in A$, since if not, A with $x \notin A \in \mathcal{G}^3$ is not a triangle, contradicting Claim 2. But then $|\mathcal{G}^3(x)| \geqslant 4$, $|\mathcal{G}^3 - x| \geqslant \tau(\mathcal{G}^3 - x) \geqslant 2$, contradicting the assumption $5 \geqslant |\mathcal{G}^3| (\geqslant 4 + 2 = 6)$.

https://doi.org/10.1017/S096354830800922X Published online by Cambridge University Press

It follows that A meets one of T_3 and T_4 and not only in z. Indeed,

$$|A \setminus \{z\}| + |T_3 \setminus \{z\}| + |T_4 \setminus \{z\}| \geqslant 3 + 2 + 2 = 7 > 6 = |[n] \setminus \{z\}|.$$

Let this element be $x \in A \cap T_3 \setminus \{z\}$; since $x \in T_1 \cup T_2$, we can assume, for instance, $x \in T_1$.

Now, again, Claim 2 can be applied to x, and since $|\mathcal{G}^3| \leqslant 5$, both triangles of $\mathcal{G} - x$ are already among the listed sets. These can be only T_2 and T_4; in particular, T_4 is also a triangle.

So $T_4 = (T_1 \setminus \{x\}) \cup \{z\}$. Because of Claim 3, T_3 also contains, besides $x \in T_1$, an element of T_2. Finally, $A = \{x, z\} \cup (T_2 \setminus T_3)$, since any other element in A would again contradict Claim 2. It follows that $\mathcal{G}^4 = \{A\}$. In order to 2-generate $\{1, 2, 3, 4, 5, 6, 7\}$ itself, we need a set in $\mathcal{G}^4$ and its complement. But the complement of A is different from all of T_1, T_2, T_3, T_4, so $\mathcal{G}^3$ has a sixth element, and this final contradiction finishes the proof of the theorem. □

Corollary A3. *For any* (*not necessarily optimal*) $(8, 2)$*-generator, then* (3.1) *holds, there exists* $z \in [n]$ *such that* (3.5) *holds, and* $\mathcal{Y}(8, 2)$ *is the unique optimal* $(8, 2)$*-generator.*

Proof. $|\mathcal{G}^4| \geqslant 2$ is obvious, as usually

$$\sum_{z \in [n]} |\mathcal{G}^3 - z| \geqslant 8\, Y^3(7, 2) = 48$$

(since each 7-element set still contains $g \in \mathcal{G}$, $|g| \geqslant 4$), and every set of $\mathcal{G}^3$ has been counted at most 5 times in this sum, so $|\mathcal{G}^3| \geqslant \lceil 48/5 \rceil = 10 = Y^3(8, 2)$.

It is now easy to prove $|\mathcal{G}^2| \geqslant Y^2(8, 2)$ (and similarly $|\mathcal{G}^1| \geqslant Y^1(8, 2)$), with the same argument as in the proof of the previous theorem: it suffices to prove that $\hat{\mathcal{G}}$ with $|\hat{\mathcal{G}}| = Y(8, 2) = 30$ is simply $\mathcal{Y}(8, 2)$, and for this it suffices to prove that $\hat{\mathcal{G}}$ has an element of degree $2^3 = 8$. So the only remaining assertion to prove is that for any $(8, 2)$-generator with $|\mathcal{G}| = Y(8, 2) = 30$ there exists $z \in [n]$ such that (3.5) holds. Then the last assertion also follows by Lemma 3.10. Let $\mathcal{G}$ be such an (n, k)-generator.

Let $d := 1/n \sum_{x \in [n]} \hat{\mathcal{G}}(x)$ be the *average degree* of $\mathcal{G}$. Clearly (as before, see Proposition 3.2),

$$d = 1/n \sum_{x \in [n]} |\mathcal{G}(x)| = 1/n \sum_{g \in \hat{\mathcal{G}}} |g| = 1/n \sum_{i=1}^{n} |\mathcal{G}^i| = 1/8(30 + 22 + 10 + 1) > 7,$$

finishing the proof of the corollary. □

Note that for this last statement a much weaker bound is sufficient, namely the first easy estimate of $|\mathcal{G}^3|$ without the harder one.

In $\mathcal{Y}(8, 2)$ the average degree is equal to the maximum degree and the same could be proved for the optimum generator, which is why Corollary A3 includes Conjecture 2.3. The same can be proved for arbitrary even n and $k = 2$, but the odd n case with 'small' average degree remains open.

https://doi.org/10.1017/S096354830800922X Published online by Cambridge University Press

Acknowledgement

We are indebted to Nicolas Trotignon for useful discussions, particularly for noticing the variety of optimal generators when $p = 2$. We also thank Zoltán Füredi for connecting us to the current state of the subject.

References

[1] Da Cunha, C. (2004) Definition and inventory management of semi-finished products in an Assembly To Order context. PhD Thesis, INPG, Grenoble. (In French.)
[2] Erdős, P. (1993) Private communication mentioned in [3].
[3] Füredi, Z. and Katona, G. O. H. (2006) 2-bases of quadruples. *Combin. Probab. Comput.* **15** 131–141.
[4] Garey, M. R. and Johnson, D. S. (1979) *Computers and Intractability: A Guide to the Theory of NP-Completeness*, Freeman.
[5] Sidorenko, A. F. (1995) What we know and what we do not know about Turán numbers. *Graphs Combin.* **11** 179–199.
[6] Turán, P. (1941) On an extremal problem in graph theory. *Math. Fiz. Lapok* **48** 436–452. (In Hungarian.)

https://doi.org/10.1017/S096354830800922X Published online by Cambridge University Press