Available online at www.sciencedirect.com
ScienceDirect
Electronic Notes in Discrete Mathematics 29 (2007) 73–77
Electronic Notes in DISCRETE MATHEMATICS
www.elsevier.com/locate/endm

Optimizing diversity [1]

Yannick Frein [2], Benjamin Lévêque [3], András Sebő [4]

Laboratoire G-SCOP, INPG, UJF, CNRS
46, avenue Felix Viallet, 38031 Grenoble Cedex, France

Abstract

We consider the problem of minimizing the size of a set system $\mathcal{G}$ such that every subset of $\{1, \ldots, n\}$ can be written as a disjoint union of at most k members of $\mathcal{G}$, where k and n are given numbers. This problem is originating in a real-world application aiming at the diversity of industrial production, and at the same time the $k = 2$ case is a question of Erdős, studied recently by Füredi and Katona. We conjecture that a simple construction providing a feasible solution is optimal for this problem; we prove this conjecture in special cases, complementary to those solved by Füredi and Katona, in particular for the case $n \leq 3k$. These special cases occur to be interesting from the viewpoint of the application as well.

Keywords: Turán type problems, extremal problems in graphs and hypergraphs, diversity, semi-finished products.

[1] This research has been supported by the ADONET network of the European Community, which is a Marie Curie Training Network.

[2] Email: `yannick.frein@g-scop.inpg.fr`
[3] Email: `benjamin.leveque@g-scop.inpg.fr`
[4] Email: `andras.sebo@g-scop.inpg.fr`

doi:10.1016/j.endm.2007.07.013

1 Introduction

The n-element set $\{1, \ldots, n\}$ is denoted by $[n]$. For two positive integers n, k, a set system $\mathcal{G}$ is called a generator of $[n]$ if every non empty subset of $[n]$ can be obtained as a disjoint union of at most k members of $\mathcal{G}$. The problem we are studying in this paper is to find a generator of minimum size. Then we say that $\mathcal{G}$ is *optimal*, and introduce the notation $\text{opt}(n,k) := |\mathcal{G}|$. Note that every generator contains the n subsets of size 1, so $\text{opt}(n,k) \geq n$. A generator can be represented by a hypergraph where the vertices are the elements of $[n]$ and the hyperedges the members of $\mathcal{G}$.

As Zoltán Füredi reports [3], Paul Erdős [2] asked about the case $k = 2$ allowing the target-sets to be *non-disjoint* unions of the two generators. 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 change when the disjointness of the two sets is required.

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 $s \leq 2$ the problem is void, and for $s = 3$ the solution immediately follows from Turán's theorem [5]. For $s = 4$ and $n \geq 8$ they establish that the cardinality of an optimal generator is $n + \binom{n}{2} - \lfloor \frac{4}{3}n \rfloor$. When $s \leq 4$ it does clearly not matter whether the two sets are required to be disjoint or not.

These problems have been independently studied for optimizing the diversity of industrial production. To answer market requirements, many companies, namely in the motorcar industry, want to reduce the delay between the command 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 generator for the problem (n,k), where n is the number of options, and k the maximum number of semi-finished products that can be assembled. Hard to believe, but from the viewpoint of industrial technology neither the *disjointness* constraint nor the exact generation of *all subsets* can be relieved. The combinatorial optimization problem where the sets to be generated are given as input is NP-Hard anyway, so in view of generating all sets, it is better to look at the problem with the viewpoint of extremal combinatorics. A confirmation of this approach is that experts in production management [1] have arrived at the same conjecture as Erdős. However, the only result about this problem seems to be [3].

2 Construction

A natural way of constructing 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 such that $n = p\,k - r$ with $0 \le r < k$. 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:

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

where $\mathcal{P}(V)$ is the power set of V. The cardinality of such a generator is $\mathrm{constr}(n,k) := r \times (2^{p-1} - 1) + (k-r) \times (2^p - 1)$. Note that

$$\mathrm{constr}(n,k) = \mathrm{constr}(n-1,k) + 2^{p-1},$$

and it is this simple recursive formula that seems to be useful to keep in mind. Our goal is to prove *the same recursive formula for* $\mathrm{opt}(n,k)$.

The example of figure 1 shows that for $n = 13$ and $k = 5$, we have $\mathrm{constr}(13,5) = 27$. There are 13 members of size 1, 11 members of size 2 and 3 members of size 3.

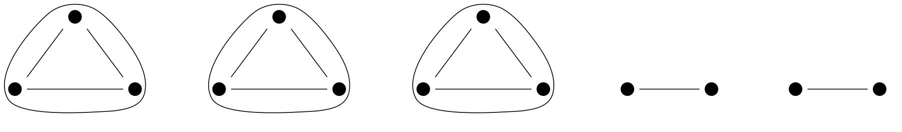

Fig. 1. Hypergraph representation for $n = 13$, $k = 5$

The construction gives an upper bound on the cardinality of an optimal generator : $\mathrm{opt}(n,k) \le \mathrm{constr}(n,k)$. We conjecture that this inequality is in fact an equality.

Conjecture 1 *For all* $n, k \in \mathbb{N}$, $\mathrm{CONSTR}(n,k)$ *is an optimal generator.*

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

Indeed, as Zoltán Füredi report [3], Erdős [2] asked the same question for $k = 2$ without requiring the disjointness of the sets. Could the same assertion be true for arbitrary k ? Let $\mathrm{opt}'(n, k)$ denote the optimum for this problem. Clearly, $\mathrm{opt}'(n, k) \leq \mathrm{opt}(n, k) \leq \mathrm{constr}(n, k)$, so if Erdős's conjecture is true for arbitrary k there is equality throughout. This would mean that disjointness is an irrelevant requirement. (In the sense that it does not change the optimum value.) Could this be proved ?

Moreover, we also conjecture unicity of the construction.

Conjecture 2 *For every $n, k \in \mathbb{N}$ such that $p(n, k) \neq 2$,* $\mathrm{CONSTR}(n, k)$ *is the unique optimal generator.*

In order to obtain optimality and unicity results for the construction, we introduce the following conjecture.

Conjecture 3 *Every generator contains a vertex of degree at least 2^{p-1}.*

We prove that Conjecture 3 implies both Conjectures 1 and 2 and that Conjecture 3 is true for $p = 1, 2, 3$. This implies results on optimality and unicity that are summarized in Figure 2.

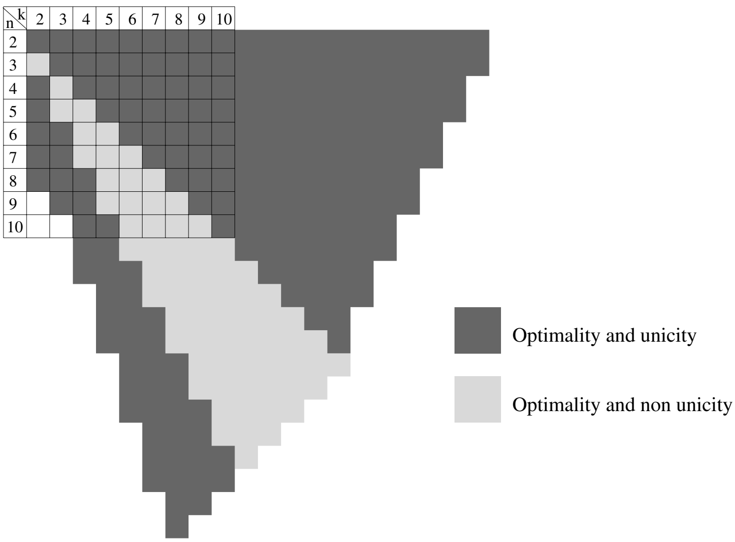

Fig. 2. Results of optimality and unicity of the construction

Notice that the partition underlying the construction is the same as that in Turán's theorem [5]. The two are actually related. The Turán number $T(n, s, l)$, where n, s, l are three positive integers with $l \leq s \leq n$, 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 when $l = 2$. 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. Turán proved that the unique optimum is given by k cliques of almost equal size that partition the vertex-set. This partition coincides with the defining partition of the construction ! This shows that, the number of members of size at least two in a generator is at least the number of members of size exactly two of the construction.

When $l \geq 3$, Turán conjectured that the partition into blocks still gives the solution to its problem but this appears to be false. According to Sidorenko [4], when $n = 9$, $s = 5$, $l = 3$ with $s = (l-1)k+1$ and $k = 2$, Turán's construction provides $\binom{4}{3} + \binom{5}{3} = 14$ subsets of size 3 whereas the affine plane of order 3 gives a solution with only 12 subsets. The solution of Füredi and Katona [3] is a modification and generalisation of this affine plane.

We cannot continue in this direction since finding the Turán number when $l \geq 3$ is known as a difficult open problem, moreover a closer direct look using more than just the containments provides better estimates for the diversity problem.

References

[1] Da Cunha, C., "Definition and inventory management of semi-finished products in an Assembly To Order context", PhD Thesis, INPG, Grenoble, 2004.

[2] Erdős, P., personal communication mentioned in [3], 1993.

[3] Füredi, Z., and G.O.H. Katona, *2-bases of quadruples*, Combinatorics, Computing and Probabality **15** (2006), 131-141.

[4] Sidorenko, A.F., *What we know and what we do not know about Turán numbers*, Graphs Combin. **11** (1995), 179-199.

[5] Turán, P., *On an extremal problem in graph theory*, Math. Fiz. Lapok **48** (1941), 436-452.