Combinatorics, Probability and Computing (2011) 20, 743–747.
doi:10.1017/S0963548311000319

Daisies and Other Turán Problems

BÉLA BOLLOBÁS[1,2†], IMRE LEADER[2] and CLAUDIA MALVENUTO[3]

[1]Department of Mathematical Sciences, University of Memphis, Memphis TN 38152, USA
(e-mail: b.bollobas@dpmms.cam.ac.uk)

[2]Department of Pure Mathematics and Mathematical Statistics, Wilberforce Road, Cambridge CB3 0WB, UK
(e-mail: i.leader@dpmms.cam.ac.uk)

[3] Dipartimento di Informatica, Sapienza – Università di Roma, Via Salaria 113, 00198 Roma, Italy
(e-mail: claudia@di.uniroma1.it)

Received 8 May 2011; revised 9 July 2011

Our aim in this note is to make some conjectures about extremal densities of daisy-free families, where a 'daisy' is a certain hypergraph. These questions turn out to be related to some Turán problems in the hypercube, but they are also natural in their own right. We start by giving the daisy conjectures, and some related problems, and shall then go on to describe the connection with vertex-Turán problems in the hypercube.

This note is self-contained. Our notation is standard: in particular, we write $[n]$ for $\{1,\ldots,n\}$, and Q_n for the n-dimensional hypercube (the set of all subsets of an n-point set). For a set X, we write $X^{(r)}$ for the set of all r-element subsets of X. An r-graph (or r-uniform hypergraph) on X is a subset of $X^{(r)}$. For background on hypergraphs see [2], and for background on Turán problems in general see [10] and [8].

A *daisy*, or *r-daisy*, is a certain r-uniform hypergraph consisting of six sets: given an $(r-2)$-set P and a 4-set Q disjoint from P, the daisy on (P,Q) consists of the r-sets A with $P \subset A \subset P \cup Q$. We write this as $\mathcal{D}$, or $\mathcal{D}_r$. Our fundamental question is: How large can a family $\mathcal{A}$ of r-sets from an n-set be if $\mathcal{A}$ does not contain a daisy?

As usual, if $\mathcal{F}$ is a family of r-sets, we write $\mathrm{ex}(n,\mathcal{F})$ for the maximum size of a family of r-sets from an n-set that does not contain a copy of $\mathcal{F}$, and $\pi(\mathcal{F})$ or $\pi_r(\mathcal{F})$ for the limiting density, namely the limit of $\mathrm{ex}(n,\mathcal{F})/\binom{n}{r}$ as n tends to infinity; a standard averaging argument shows that this limit exists, and indeed that $\mathrm{ex}(n,\mathcal{F})/\binom{n}{r}$ is a decreasing function of n.

† Research supported in part by NSF grants CNS-0721983, CCF-0728928 and DMS-0906634, and ARO grant W911NF-06-1-0076.

Conjecture 1. *$\pi(\mathcal{D}_r) \to 0$ as $r \to \infty$.*

What is unusual here is that we are not so concerned with the actual values of $\pi_r(\mathcal{D}_r)$ for particular r: our main interest is in the *limit* of these values. We will see later why Conjecture 1 is related to Turán questions in the hypercube.

Since the hypergraph $\mathcal{D}_r$ is not r-partite, it follows that $\pi(\mathcal{D}_r) \geqslant r!/r^r$, as the complete r-partite r-graph does not contain a daisy. For $r = 2$, a daisy is precisely a K_4, and so Turán's theorem tells us that $\pi(\mathcal{D}_2) = 2/3$. Although, even for $r = 3$, we do not know what the limiting density is, we believe we know what it should be.

Conjecture 2. $\pi(\mathcal{D}_3) = 1/2$.

To see where this conjecture comes from, note that the 3-graph on 7 vertices given by the complement of the Fano plane does not contain a daisy. Here, as usual, the Fano plane is the projective plane over the field of order 2; equivalently, it consists of the triples $\{a, a+1, a+3\}$, where the ground set is the integers mod 7. This gives $\mathrm{ex}(7, \mathcal{D}_3) \geqslant 28 = \frac{4}{5}\binom{7}{3}$. If we take a blow-up of this, thus dividing $[n]$ into 7 classes $C_0, \ldots, C_6$ each of size $\lfloor n/7 \rfloor$ or $\lceil n/7 \rceil$ and taking the 7-partite 3-graph consisting of all 3-sets whose 3 classes are not $\{C_a, C_{a+1}, C_{a+3}\}$ (with subscripts taken mod 7), we obtain $\mathrm{ex}(n, \mathcal{D}_3) \geqslant (1 + o(1))\frac{24}{49}\binom{n}{3}$. But now we may iterate, taking a similar construction inside each class, and so on. This gives a limiting density of $24/49$ times $1 + 1/49 + 1/49^2 + \cdots$, which is exactly $1/2$.

We do not even see any counter-example to a much stronger assertion, that this is the actual best-possible example, at least if n is a power of 7. This reduces to the following conjecture.

Conjecture 3. *Let $n = 7^k$, and let $\mathcal{A}$ be a family of 3-sets of $[n]$ not containing a daisy. Then $|\mathcal{A}| \leqslant (1 - 1/49^k)\, n^3/12 = \frac{1}{2}\binom{n+1}{3}$.*

This conjecture was independently made by Goldwasser [5], based on the same Fano plane construction.

The above 'daisy' is actually part of a more general family. In general, an (s,t)-*daisy* $\mathcal{D}(s,t) = \mathcal{D}_r(s,t)$ consists of all of those r-sets A that contain a fixed $(r-t)$-set P and are contained in $P \cup Q$, where Q is a fixed s-set disjoint from P. Thus a $(4,2)$-daisy is precisely a daisy in our earlier sense.

Conjecture 4. *Let s and t be fixed. Then $\pi(\mathcal{D}_r(s,t)) \to 0$ as $r \to \infty$.*

Perhaps the most natural case of this is when $s = 2t$: see later. In fact, in a sense this is the *only* case: since $\mathcal{D}_r(s,t)$ is contained in $\mathcal{D}_r(s+1,t)$ and also in $\mathcal{D}_r(s+1,t+1)$, to verify Conjecture 4 it would be enough to verify it for the case $s = 2t$.

Conjecture 4 certainly holds when $t = s-1$, as then we are simply asking that our family should contain no s r-sets from any $(r+1)$-set. Averaging gives $\pi(\mathcal{D}_r(s,s-1)) \leqslant \frac{s-1}{r+1}$, which tends to zero as required. Conjecture 4 also holds if $t = 1$, as our condition is now that no $(r-1)$-set can be contained in s r-sets in our family. Hence our family has size at most $\binom{n}{r-1}(s-1)/r$,

https://doi.org/10.1017/S0963548311000319 Published online by Cambridge University Press

whence $\pi(\mathcal{D}_r(s,1)) = 0$ for all r. (Alternatively, as $\mathcal{D}_r(s,1)$ is r-partite, one may use the well-known result of Erdős [4] that the limiting density for any r-partite r-graph is zero.) Thus our starting case of the $(4,2)$-daisy is in fact the first non-trivial case.

We digress briefly to point out that a related notion is far simpler to analyse. A daisy (a $(4,2)$-daisy) consists of 6 r-sets in a set of size $r+2$. Suppose that, rather than forbidding an actual daisy, we instead do not allow any $(r+2)$-set to contain 6 (or more) r-sets. In this case it is easy to see that we cannot have a constant proportion of the r-sets (as $r \to \infty$), because averaging gives that the proportion of r-sets in our family is at most $6/\binom{r+2}{2}$.

The situation is the same if we replace our '6' with any function that is $o(r^2)$. However, this changes the moment we reach a constant times r^2. Indeed, suppose that we wish to insist that no $(r+2)$-set contains cr^2 r-sets, for a given constant c. Partition $[n]$ into k sets of size n/k (for some suitable constant k), and take the family $\mathcal{A}$ of all r-sets that have between $(1-\delta)r/k$ and $(1+\delta)r/k$ points in each class (for some small constant δ) and have even-size intersection with each class (or, if r is odd, one intersection size is odd). This is a positive proportion of all r-sets, and yet no $(r+2)$-set R can contain cr^2 sets from $\mathcal{A}$. Indeed, R would have to meet every class of the partition in roughly between $(1-\delta)r/k$ and $(1+\delta)r/k$ points (or else it will contain no sets from $\mathcal{A}$). And now it is easy to check that if R meets all classes in an even number of points then the number of sets of $\mathcal{A}$ contained in R is less than cr^2 (if k is large enough), and similarly if the intersection sizes of R with the classes have any given parities.

Let us remark that the notion of an (s,t)-daisy is only the 'tip of the iceberg'. Indeed, more generally we could combine any two hypergraphs, in the sense that we combined one $(r-t)$-set and the family of all t-sets from an s-set to form the (s,t)-daisy. Thus, given hypergraphs $\mathcal{F}$ and $\mathcal{G}$, we define $\mathcal{F} * \mathcal{G}$ to be the hypergraph, on ground-set the disjoint union of the ground-sets of $\mathcal{F}$ and $\mathcal{G}$, whose edges are all sets of the form $A \cup B$, where $A \in \mathcal{F}$ and $B \in \mathcal{G}$. For example, if both $\mathcal{F}$ and $\mathcal{G}$ are complete graphs, say on s and t points respectively, then $\mathcal{F} * \mathcal{G}$ is a 4-graph consisting of all 4-sets on $[s+t]$ that meet $[s]$ in exactly 2 points.

A rather general question is as follows.

Problem 5. *Let $\mathcal{F}$ be an r-graph and $\mathcal{G}$ be an s-graph. How does $\pi_{r+s}(\mathcal{F} * \mathcal{G})$ compare to $\pi_r(\mathcal{F})$ and $\pi_s(\mathcal{G})$?*

One very interesting case of this is when $\mathcal{F}$ and $\mathcal{G}$ are the same hypergraph. More generally, let us write $\mathcal{F}^d$ for the d-fold product $\mathcal{F} * \ldots * \mathcal{F}$.

Problem 6. *Let $\mathcal{F}$ be a fixed r-graph. As d varies, how does $\pi_{dr}(\mathcal{F}^d)$ behave?*

We do not even know what happens when $\mathcal{F} = [s]^{(r)}$, *i.e.*, $\mathcal{F}$ consists of all r-sets of an s-set.

We now turn to the connection with Turán problems in the hypercube. Indeed, it was this link that led us to define the notion of a daisy in the first place. The basic vertex-Turán problem in the hypercube Q_n is as follows: How many points do we need to meet all the d-cubes of an n-cube? We are interested in the behaviour as n gets large, for fixed d. (We mention in passing that there are also a host of edge-Turán problems in the hypercube: see [1] and the references therein.)

We clearly need at least a fraction $1/2^d$ (of the total number of points, 2^n), just to meet all of the d-cubes in a given direction. From the other side, if we take every $(d+1)$st layer of the

https://doi.org/10.1017/S0963548311000319 Published online by Cambridge University Press

n-cube (where a layer means $[n]^{(r)}$ for some r) then we certainly meet every d-cube, and this shows that we can take a fraction $1/(d+1)$ of the n-cube.

Let us write t_d for the limiting density (which exists, by averaging). The behaviour of t_d was investigated by Alon, Krech and Szabó [1], who showed that in a $(d+2)$-cube we need at least $\log d$ points to meet every d-cube (logs are to base 2). By averaging, this gives that t_d is at least $(\log d)/2^{d+2}$. And, remarkably, these bounds of $(\log d)/2^{d+2} \leqslant t_d \leqslant 1/(d+1)$ are all that is known in general about the asymptotic behaviour of t_d. The only exact values that are known are t_1, which is trivially seen to be $1/2$, and t_2, which is $1/3$, as shown by E. A. Kostochka [9] and by Johnson and Entringer [7]. See also Johnson and Talbot [6] for related results.

We believe that $t_d = 1/(d+1)$, and, as we now explain, the problems on daisies relate to this.

Suppose we consider the case $d = 4$ (it turns out to be slightly simpler to consider d even), and we look at just those 4-cubes that go from layer $\frac{n}{2} - 2$ to layer $\frac{n}{2} + 2$ (assuming that n is even): we call these the *middle* 4-cubes. And suppose further that we wish to meet all of these cubes using only points in the middle layer of the cube. We conjecture that nearly all of the points of the middle layer must be used.

Conjecture 7. *Let n be even, and let $\mathcal{A}$ be a subset of $[n]^{(n/2)}$ that meets every middle 4-cube. Then $|\mathcal{A}| \geqslant (1 - o(1))\binom{n}{n/2}$.*

We think that Conjecture 7 might be the ‘right first step’ in showing that $t_4 = 1/5$.

We claim that Conjecture 1 implies Conjecture 7. Indeed, suppose that (for n large) $\mathcal{A}$ is a subset of $[n]^{(n/2)}$ that meets every middle 4-cube. For a given value of r, consider those sets in $\mathcal{A}$ that contain a fixed $(\frac{n}{2} - r)$-set R: this corresponds exactly to a family of r-sets (from a ground-set of size $\frac{n}{2} + r$) that meets every daisy, and so by Conjecture 1 has size at least $(1 - o(1))\binom{\frac{n}{2}+r}{r}$. Averaging over all such R, we obtain $|\mathcal{A}| \geqslant (1 - o(1))\binom{n}{n/2}$, as required.

In fact, Conjecture 1 is actually equivalent to Conjecture 7. For Conjecture 7, in the language of daisies, states precisely that $\mathrm{ex}(n, \mathcal{D}_{n/2})/\binom{n}{n/2} \to 0$ as $n \to \infty$, which implies that $\pi(\mathcal{D}_{n/2}) \to 0$.

Similarly, we make the following conjecture, which we hope would be a step towards showing that $t_d = 1/(d+1)$.

Conjecture 8. *Let d be fixed. Let n be even, and let $\mathcal{A}$ be a subset of $[n]^{(n/2)}$ that meets every middle $2d$-cube. Then $|\mathcal{A}| \geqslant (1 - o(1))\binom{n}{n/2}$.*

Just as Conjecture 1 is equivalent to Conjecture 7, so Conjecture 4 for the parameters $(2d, d)$ is equivalent to Conjecture 8. This is why the case $s = 2t$ seems the most interesting case of Conjecture 4.

Finally, we mention briefly a beautiful conjecture of Johnson and Talbot [6], about meeting d-cubes in several points, that is also closely tied to our daisy problems. They conjecture that if we have a positive fraction of the vertices of the n-cube then (for n sufficiently large) there must be some d-cube containing at least $\binom{d}{\lfloor d/2 \rfloor}$ points of our family. (This is the greatest number of points of a d-cube that one could ask for, because of the family consisting of every $(d+1)$st layer of the n-cube.)

It is easy to see that Conjecture 4 is actually equivalent to this conjecture. Indeed, if $\mathcal{A}$ is a subset of Q_n of positive density then $\mathcal{A}$ must contain a positive proportion of a layer not far from the middle layer of the n-cube, and Conjecture 4 (plus averaging) now yields a $\mathcal{D}_r(d, \lfloor d/2 \rfloor)$ for suitable r just as above. In the other direction, if Conjecture 4 were false then, by putting together suitable counter-examples on every $(d+1)$st layer (for layers not far from the middle layer), we could find a subset of the n-cube of positive density that did not contain $\binom{d}{\lfloor d/2 \rfloor}$ points of any d-cube.

This connection with the Johnson–Talbot conjecture was independently observed by Bukh [3], who also made Conjecture 4 independently.

References

[1] Alon, N., Krech, A. and Szabó, T. (2007) Turán's theorem in the hypercube. *SIAM J. Discrete Math.* **21** 66–72.

[2] Bollobás, B. (1986) *Combinatorics*, Cambridge University Press.

[3] Bukh, B. (2011) Personal communication.

[4] Erdős, P. (1964) On extremal problems of graphs and generalized graphs. *Israel J. Math.* **2** 459–464.

[5] Goldwasser, J. (2011) Personal communication.

[6] Johnson, J. R. and Talbot, J. (2010) Vertex Turán problems in the hypercube. *J. Combin. Theory Ser. A* **117** 454–465.

[7] Johnson, K. A. and Entringer, R. (1989) Largest induced subgraphs of the n-cube that contain no 4-cycles. *J. Combin. Theory Ser. B* **46** 346–355.

[8] Keevash, P. Hypergraph Turán problems. In *Surveys in Combinatorics 2011*, Cambridge University Press, pp. 83–139.

[9] Kostochka, E. A. (1976) Piercing the edges of the n-dimensional unit cube (in Russian). *Diskret. Analiz Vyp. 28 Metody Diskretnogo Analiza v Teorii Grafov i Logiceskih Funkcii* 55–64.

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