FrontierMath Open Problems

Epoch AI

A Ramsey-style Problem on Hypergraphs

A precise statement of the question.

A *hypergraph* is a set V (called *vertices*) together with a collection $\mathcal{H}$ of subsets of V (called *hyperedges*). A hypergraph $(V, \mathcal{H})$ is said to *contain a partition of size* n if there is some $D \subseteq V$ and $\mathcal{P} \subseteq \mathcal{H}$ such that $|D| = n$ and every member of D is contained in exactly one member of $\mathcal{P}$. In this case, we say that D and $\mathcal{P}$ form a partition in $(V, \mathcal{H})$.

For example, the picture on the left above shows a hypergraph with 8 vertices and 4 hyperedges. The picture on the right shows that it contains a partition of size 4.

Given a hypergraph $(V, \mathcal{H})$, a vertex $v \in V$ is called *isolated* if $v \notin \bigcup \mathcal{H}$.

Question. *Find an example of a hypergraph* $(V, \mathcal{H})$ *with no isolated vertices such that* $|V| \geq 66$*, and* $(V, \mathcal{H})$ *contains no partitions of size* > 20*.*

The format of the answer should be as follows. Without loss of generality, we may (and do) assume the vertices of the graph are positive whole numbers, for example $V = \{1, 2, 3, \ldots, 66\}$. The members of $\mathcal{H}$ are subsets of V, and should be expressed as a list of sets separated by commas. The sets are represented as lists of numbers separated by commas and encased in curly braces. For example, the hypergraph illustrated above, after labelling the vertices appropriately, is expressed in standard set-theoretic notation as:

$$\Big(\{1, 2, 3, 4, 5, 6, 7, 8\}, \big\{\{1, 3, 4, 5\}, \{2, 3, 4, 5\}, \{4, 5, 6, 7\}, \{4, 5, 6, 8\}\big\}\Big).$$

The appropriate way to format this hypergraph as an answer would be:

```
{1,3,4,5},{2,3,4,5},{4,5,6,7},{4,5,6,8}
```

Note that there is no need to input the list of vertices: it is redundant because if a hypergraph $(V, \mathcal{H})$ has no isolated points, then $V = \bigcup \mathcal{H}$.

Background.

The idea for this question is taken from the paper

> W. Brian and P. B. Larson, "Choosing between incompatible ideals," *European Journal of Combinatorics* **26** (2021), article no. 103349.

This paper explores several Ramsey-theoretic problems arising from studying the simultaneous convergence of sets of infinite series. One of these problems centers around a sequence of numbers, which are denoted $H(1), H(2), H(3), \ldots,$ and defined as follows.

Definition 1. *Define $H(n)$ to be the greatest $k \in \mathbb{N}$ such that there is a hypergraph $(V, \mathcal{H})$ with $|V| = k$ having no isolated vertices and containing no partitions of size greater than n.*

In this terminology, our question is simply: *Show* $H(20) \geq 66$.

It is not immediately clear that the number $H(n)$ is well-defined for all n. *A priori*, it seems that there could be arbitrarily large hypergraphs having no isolated vertices and containing no partitions of size greater than n. Part of the paper is devoted to showing this is not the case, so $H(n)$ is always well-defined, and to putting reasonable bounds on $H(n)$. The bounds given in the paper are roughly:

$$\tfrac{1}{2} n \log_2(n) - \tfrac{1}{2} n \; < \; H(n) \; < \; n \ln n + \gamma n,$$

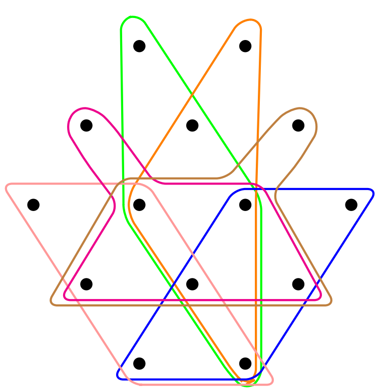
where γ is the Euler-Mascheroni constant. More precisely, the best known upper bound is $H(n) \leq \sum_{k=1}^{n} \frac{n}{k}$, and the lower bound can be expressed as $k_n \leq H(n)$, where the sequence $k_1, k+_2, k_3, \ldots$ is defined recursively by the formula

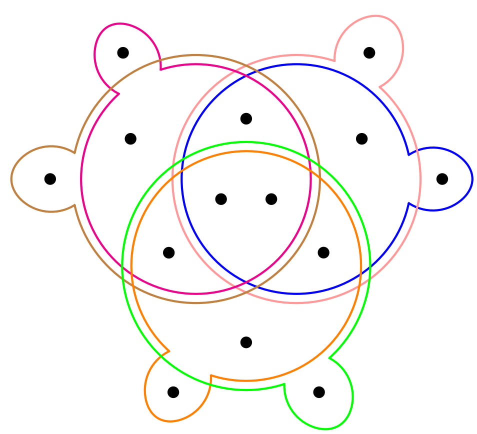
$$k_1 = 1 \qquad \text{and} \qquad k_n = \left\lfloor \frac{n}{2} \right\rfloor + k_{\lfloor \frac{n}{2} \rfloor} + k_{\lfloor \frac{n+1}{2} \rfloor}.$$

The precise values in the sequence $H(1), H(2), H(3), \ldots$ have proved very difficult to pin down. Only the smallest few values are known. The sequence begins:

$$H(1) = 1, \; H(2) = 3, \; H(3) = 5, \; H(4) = 8, \; H(5) = 10, \; H(6) = 14, \; H(7) = 17.$$

The values of $H(1)$ through $H(5)$ follow immediately from the bounds given above, since (after rounding the upper bound down to the nearest integer) these bounds match for $n = 1, 2, 3, 4, 5$. The bounds above give $13 \leq H(6) \leq 14$. Showing that $H(6) = 14$ requires finding a hypergraph of size 14 that contains no partitions of size > 6. Here are two such examples:





One thing to notice is that while these examples were rather difficult to find, it is very easy to verify that they are examples. Indeed, one can check by hand, in just a few minutes, that these hypergraphs do not contain any partitions of size > 6.

The fact that $H(7) = 17$ is the result of a direct computer search carried out as the honors project of an undergraduate student, Lauren Ostrea, in 2022. She wrote a Java program that systematically searched for hypergraphs of size 17 and 18 with no partitions of size > 7. Her program had to run overnight to verify (by exhaustion) that there is no example of size 18. Even though the size-17 example took an enormous amount of computation to find by brute force, verifying that it had the required properties was completely straightforward. For hypergraphs of this size, it is still a feasible task to do by hand.

Why such a hypergraph should exist.

Jacob Page, a master's student at UNC Charlotte, wrote his thesis on the problem of finding improved bounds for the $H(n)$:

J. D. Page, "Improved bounds on a combinatorial problem," *Master's thesis*, University of North Carolina at Charlotte, 2021.

While he was not able to improve our upper bounds, he was able to find a different approach to obtaining lower bounds. This approach starts with an intricate construction of hypergraphs with no large partitions. His approach yields better lower bounds for most values of n, but not all. Asymptotically his bounds are the same, and constitute (roughly) an improvement on the original $\approx n \log_2 n$ bound by a term of order $\log n$.

In addition to the student project by Lauren Ostrea and the master's project by Jacob page, several other mathematicians have taken an interest in the sequence $H(n)$. The main reason for this interest is that the sequence is intimately related to problems from infinite combinatorics concerning the simultaneous convergence of many infinite series. These problems on infinite series were in fact the original motivation for the Brian-Larson paper. The problems arise in:

J. Brendle, W. Brian, and J. D. Hamkins, "The subseries number," *Fundamenta Mathematicae* **247** (2019), pp. 49–81.

W. Brian, "Three conditionally convergent series, but not four," *Bulletin of the London Mathematical Society* **51** (2019), pp. 207–222.

I have corresponded further with Paul Larson about this problem, on and off since we published our paper in 2021. Jörg Brendle has also retained an interest in the related infinite series problems, and worked on ideas close to the $H(n)$ sequence since 2019. Tristan van der Vlugt has also taken an interest in this constellation of questions, and he and I have corresponded about them recently.

The consensus seems to be that improving the bounds on the $H(n)$, in either direction, will require some kind of conceptual breakthrough. A brute force approach seems impossible (as described below), but the work of Page suggests that the lower bounds can be improved by finding clever constructions of hypergraphs.

Indeed, we all suspect that the lower bounds may be more susceptible to improvement than the upper bounds. This suspicion arises in part from the fact that the current upper bounds are much more difficult to prove than the current lower bounds – the proof of the latter feels somewhat coarse and inefficient. Furthermore, in the first two cases where the lower and upper bounds do not match, $n = 6$ and 7, the lower bounds have proven not to be optimal, and the work of Page shows the lower bounds from the Brian-Larson paper are not optimal for most n.

The best known lower bound on $H(20)$ is 64, and our question asks to improve this to 66. The value seems a safe compromise between going too low, and risking making the question too easy, and going too high, and risking that there is no answer.

The infeasibility of finding it by brute force.

To prove $k \leq H(n)$ for some k, one must construct a k-sized hypergraph with no partitions of size $> n$. The search space for such hypergraphs grows very rapidly with k and n. A reasonably smart search should utilize a few not-too-difficult facts:

- A hypergraph $(V, \mathcal{H})$ witnessing $H(n) \geq k$ need have no more than k vertices; in other words, there is no advantage to considering hypergraphs of size $> k$.
- A hypergraph $(V, \mathcal{H})$ witnessing $H(n) \geq k$ need have no more than n hyperedges; so we may assume, without loss of generality, that $|\mathcal{H}| \leq n$.
- A hypergraph $(V, \mathcal{H})$ witnessing $H(n) \geq k$ can have no edges of size $> n$ (because if $E \in \mathcal{H}$ and $|E| > n$, then $\mathcal{P} = \{E\}$ is a partition of size $|E| > n$).

In addition to these facts, a smart seach should also utilize the following heuristic:

- A hypergraph $(V, \mathcal{H})$ witnessing $H(n) \geq k$ should have all hyperedges $E \in \mathcal{H}$ with $|E| \approx n$.

This is true for all known examples of critical hypergraphs for the $H(n)$. Moreover, experimenting with building these hypergraphs simply gives one the feeling that having many small hyperedges is not optimal.

Taking these things into account, we can compute the number of potential hypergraphs to consider in a "smart" search for a witness to $H(n) \geq k$. There are $\binom{k}{\ell}$ possible edges of size ℓ, and (using the last heuristic), we take ℓ to go from $\frac{1}{2}n$ up to n. This gives a total of $\sum_{\ell=\frac{1}{2}n}^{n}\binom{k}{\ell}$ possible edges. We then must choose n of these possible edges to form our hypergraph. The total number of hypergraphs to be considered in this kind of search is therefore

$$\binom{\sum_{\ell=\frac{1}{2}n}^{n}\binom{k}{\ell}}{n} \approx \binom{k^n}{n} \approx k^{n^2}.$$

Being more optimistic about the search space, let us point out that having all hyperedges of size n or $n-1$ seems to be the norm for small hypergraphs witnessing $H(n) \geq k$. Making this concession (though there is no real reason to think it continues to hold for larger n), the total number of hypergraphs to be considered in this kind of search drops to $\binom{\binom{k}{n}+\binom{k}{n-1}}{n}$.

For example, even with $n = 15$ and $k = 45$, a computation on WolframAlpha gives

$$\binom{\sum_{\ell=8}^{15}\binom{45}{\ell}}{15} = \binom{627947490860}{15} \approx 7.1 \times 10^{164},$$

or for the more optimistic estimation of the search space,

$$\binom{\binom{45}{15}+\binom{45}{14}}{15} = \binom{511738760544}{15} \approx 3.3 \times 10^{163}.$$

For each candidate in this search space, one must check whether it witnesses $H(15) \geq 45$ or not. Realistically, this check requires many thousands of computations. Continuing with unwarranted optimism, however, let us simplify things by regarding each such check as a single computation. The fastest computers today perform on the order of 10^{18} computations per second. Assuming one could check 10^{18} hypergraphs per second (which, again, is wildly optimistic), it would still take on the order of 10^{138} years to check all the 3.3×10^{163} candidates in a brute force search, even when using a "smart" search and being overly optimistic about both the number of candidates to check and the time needed to check each candidate.

In short, it seems completely impossible that this kind of search could be carried out by brute force.

Natural Modifications giving rise to further questions.

A more general version of this problem is: Given some n and k, find a hypergraph witnessing $H(n) \geq k$. This problem is meaningful only in the case where k exceeds the current best known lower bound for $H(n)$, but is fairly close to it (so that it is reasonable to expect such a witness to exist). For any $n \geq 10$, the calculations above show that a brute force search for a solution should be out of reach.

The current values of n and k were chosen as a good compromise of all the following considerations:

- The number n should be large enough that it is infeasible to compute $H(n)$ by brute force. Small values of n (8 or 9) are already enough for this.
- The number n should be large enough that there is an appreciable gap between the best current lower and upper bounds for $H(n)$. This enables us to choose a number k that is more than one unit bigger than the lower bound (to avoid the possibility of an in-hindsight-trivial improvement), but not very close to the upper bound (to avoid the possibility of a witness not existing). This requires approximately $n \geq 15$.
- On the other hand, n should be small enough that a proposed solution can be checked in a short amount of time.

Listed below are the known lower and upper bounds on $H(n)$ for $n = 15, 16, \ldots, 30$. Choosing k to be a little above the best known lower bound, for any of these n values, leads to a reasonable version of the problem. We have listed two lower bounds for each n, the one on the left being the bound proved in the Brian-Larson paper, and the one on the right being the one from Jacob Page's thesis. In every case Page's bound is at least as good as the old bound.

$$
\begin{aligned}
43 &\leq 43 \leq H(15) \leq 49 \\
48 &\leq 48 \leq H(16) \leq 54 \\
50 &\leq 50 \leq H(17) \leq 58 \\
53 &\leq 56 \leq H(18) \leq 62 \\
56 &\leq 58 \leq H(19) \leq 67 \\
60 &\leq 64 \leq H(20) \leq 71 \\
63 &\leq 66 \leq H(21) \leq 76 \\
67 &\leq 72 \leq H(22) \leq 81 \\
71 &\leq 74 \leq H(23) \leq 85 \\
76 &\leq 80 \leq H(24) \leq 90 \\
79 &\leq 82 \leq H(25) \leq 95 \\
83 &\leq 88 \leq H(26) \leq 100 \\
87 &\leq 90 \leq H(27) \leq 105 \\
92 &\leq 96 \leq H(28) \leq 109 \\
96 &\leq 98 \leq H(29) \leq 114 \\
101 &\leq 104 \leq H(30) \leq 119
\end{aligned}
$$