Note on generating all subsets of a finite set with disjoint unions

David Ellis

November 2008

Abstract

We call a family $G \subset \mathcal{P}[n]$ a $k$-generator of $\mathcal{P}[n]$ if every $x \subset [n]$ can be expressed as a union of at most $k$ disjoint sets in $G$. Frein, Lévéque and Sebő [1] conjectured that for any $n \geq k$, such a family must be at least as large as the $k$-generator obtained by taking a partition of $[n]$ into classes of sizes as equal as possible, and taking the union of the power-sets of the classes. We generalize a theorem of Alon and Frankl [2] in order to show that for fixed $k$, any $k$-generator of $\mathcal{P}[n]$ must have size at least $k2^{n/k}(1 - o(1))$, thereby verifying the conjecture asymptotically for multiples of $k$.

1 Introduction

We call a family $G \subset \mathcal{P}[n]$ a $k$-generator of $\mathcal{P}[n]$ if every $x \subset [n]$ can be expressed as a union of at most $k$ disjoint sets in $G$. Frein, Lévéque and Sebő [1] conjectured that for any $n \geq k$, such a family must be at least as large as the $k$-generator

$$F_{n,k} := \bigcup_{i=1}^{k} \mathcal{P}V_i \setminus \{\emptyset\}$$

where $(V_i)$ is a partition of $[n]$ into $k$ classes of sizes as equal as possible. For $k = 2$, removing the disjointness condition yields the stronger conjecture of Erdős – namely, if $G \subset \mathcal{P}[n]$ is a family such that any subset of $[n]$ is a union (not necessarily disjoint) of at most two sets in $G$, then $G$ is at least as large as

$$F_{n,2} = \mathcal{P}V_1 \cup \mathcal{P}V_2 \setminus \{\emptyset\}$$

where $(V_1, V_2)$ is a partition of $[n]$ into two classes of sizes $\lfloor n/2 \rfloor$ and $\lceil n/2 \rceil$. We refer the reader to for example Furedi and Katona [5] for some results around the Erdős conjecture. In fact, Frein, Lévéque and Sebő [1] made the analogous conjecture for all $k$. (We call a family $G \subset \mathcal{P}[n]$ a $k$-base of $\mathcal{P}[n]$ if every $x \subset [n]$ can be expressed as a union of at most $k$ sets in $G$; they conjectured that for any $k \leq n$, any $k$-base of $\mathcal{P}[n]$ is at least as large as $F_{n,k}$.)
In this paper, we show that for \( k \) fixed, a \( k \)-generator must have size at least \( k^{2n/k}(1 - o(1)) \); when \( n \) is a multiple of \( k \), this is asymptotic to \( f(n, k) = |\mathcal{F}_{n, k}| = k(2^{n/k} - 1) \). Our main tool is a generalization of a theorem of Alon and Frankl, proved via an Erdos-Stone type result.

We first remark that for a \( k \)-generator \( G \), we have the following trivial bound on \( |G| = m \). The number of ways of choosing at most \( k \) sets in \( G \) must be at least the number of subsets of \([n]\), i.e.:

\[
\sum_{i=0}^{k} \binom{m}{i} \geq 2^n
\]

For fixed \( k \), the number of subsets of \([n]\) of size at most \( k - 1 \) is \( \sum_{i=0}^{k-1} \binom{m}{i} = \Theta(1/m) \binom{m}{k} \), so

\[
\sum_{i=0}^{k} \binom{m}{i} = (1 + \Theta(1/m)) \binom{m}{k} = (1 + \Theta(1/m))m^k/k!
\]

Hence,

\[
m \geq (k!)^{1/k} 2^{n/k}(1 - o(1))
\]

We will improve the constant from \( (k!)^{1/k} \approx k/e \) to \( k \) by showing that for any fixed \( k \in \mathbb{N} \) and \( \delta > 0 \), if \( m \geq 2^{(1/(k+1)+\delta)n} \), then any family \( G \subset \mathbb{P}[n] \) of size \( m \) contains at most

\[
\left( \frac{k!}{k^k} + o(1) \right) \binom{m}{k}
\]

unordered \( k \)-tuples \( \{A_1, \ldots, A_k\} \) of pairwise disjoint sets, where the \( o(1) \) term tends to 0 as \( m \to \infty \) for fixed \( k, \delta \). In other words, if we consider the ‘Kneser graph’ on \( \mathbb{P}[n] \), with edge set consisting of the disjoint pairs of subsets, the density of \( K_k \)'s in any sufficiently large \( G \subset \mathbb{P}[n] \) is at most \( k!/k^k + o(1) \). (This generalizes Theorem 1.3 in [2].) From the trivial bound above, any \( k \)-generator \( G \subset \mathbb{P}[n] \) has size \( m \geq 2^{n/k} \), so putting \( \delta = 1/k(k+1) \), we will see that the number of unordered \( k \)-tuples of pairwise disjoint sets in \( G \) is at most

\[
\left( \frac{k!}{k^k} + o(1) \right) \binom{m}{k}
\]

so

\[
2^n \leq \left( \frac{k!}{k^k} + o(1) + \Theta(1/m) \right) \binom{m}{k} = \left( \frac{m}{k} \right)^k (1 + o(1))
\]

and therefore

\[
m \geq k 2^{n/k}(1 - o(1))
\]

where the \( o(1) \) term tends to 0 as \( n \to \infty \) for fixed \( k \in \mathbb{N} \).
2 A preliminary Erdős-Stone type result

We will need the following generalization of the Erdős-Stone theorem:

**Theorem 1** Given \( r \leq s \in \mathbb{N} \) and \( \epsilon > 0 \), if \( n \) is sufficiently large depending on \( r, s \) and \( \epsilon \), then any graph \( G \) on \( n \) vertices with at least

\[
\frac{s(s-1)(s-2) \ldots (s-r+1)}{s^r} n^r \left( \begin{array}{c} n \\vline \kern-2.5pt \cr r \end{array} \right) + \epsilon\left(\begin{array}{c} n \\vline \kern-2.5pt \cr r \end{array}\right)
\]

\( K_r \)'s contains a copy of \( K_{s+1}(t) \), where \( t \geq C_{r,s,\epsilon} \log n \) for some constant \( C_{r,s,\epsilon} \) depending on \( r, s, \epsilon \).

Note that the density \( \eta = \eta_{r,s} := \frac{s(s-1)(s-2) \ldots (s-r+1)}{s^r} \) above is the density of \( K_r \)'s in the \( s \)-partite Turán graph with classes of size \( T \), \( K_s(T) \), when \( T \) is large.

**Proof:**

Let \( G \) be a graph with \( K_r \) density at least \( \eta + \epsilon \); let \( N \) be the number of \( l \)-subsets \( U \subset G \) such that \( G[U] \) has \( K_r \)-density at least \( \eta + \epsilon / 2 \). Then, double counting the number of times an \( l \)-subset contains a \( K_r \),

\[
N\left(\begin{array}{c} l \\vline \kern-2.5pt \cr r \end{array}\right) + \left(\begin{array}{c} n \\vline \kern-2.5pt \cr r \end{array}\right) (\eta + \epsilon / 2)\left(\begin{array}{c} l \\vline \kern-2.5pt \cr r \end{array}\right) \geq (\eta + \epsilon)\left(\begin{array}{c} n \\vline \kern-2.5pt \cr r \end{array}\right)\left(\begin{array}{c} n-r \\vline \kern-2.5pt \cr l-r \end{array}\right)
\]

so rearranging,

\[
N \geq \frac{\epsilon / 2}{1 - \eta - \epsilon / 2} \left(\begin{array}{c} n \\vline \kern-2.5pt \cr l \end{array}\right) \geq \frac{\epsilon}{2}\left(\begin{array}{c} n \\vline \kern-2.5pt \cr l \end{array}\right)
\]

Hence, there are at least \( \frac{\epsilon}{2}\left(\begin{array}{c} n \\vline \kern-2.5pt \cr l \end{array}\right) \) \( l \)-sets \( U \) such that \( G[U] \) has \( K_r \)-density at least \( \eta + \epsilon / 2 \). But Erdős proved that the number of \( K_r \)'s in a \( K_{s+1} \)-free graph on \( l \) vertices is maximized by the \( s \)-partite Turán graph on \( l \) vertices (Theorem 3 in [3]), so provided \( l \) is chosen sufficiently large, each such \( G[U] \) contains a \( K_{s+1} \).

Each \( K_{s+1} \) in \( G \) is contained in \( \left(\begin{array}{c} n-s-1 \\vline \kern-2.5pt \cr l-s-1 \end{array}\right) \) \( l \)-sets, and therefore \( G \) contains at least

\[
\frac{\epsilon}{2}\left(\begin{array}{c} n \\vline \kern-2.5pt \cr l \end{array}\right) \geq \frac{\epsilon}{2}(n/l)^{s+1}
\]

\( K_{s+1} \)'s, i.e. a positive density of \( K_{s+1} \)'s. Let \( a = s + 1 \), \( c = \frac{\epsilon}{2\log t} \) and apply the following 'blow up' theorem of Nikiforov (a slight weakening of Theorem 1 in [4]):

**Theorem 2** Let \( a \geq 2 \), \( c^a \log n \geq 1 \). Then any graph on \( n \) vertices with at least \( cn^a \) \( K_a \)'s contains a \( K_a(t) \) with \( t = \lceil c^a \log n \rceil \).

We see that provided \( n \) is sufficiently large depending on \( r, s \) and \( \epsilon \), \( G \) must contain a \( K_{s+1}(t) \) for \( t = \lceil c^{s+1} \log n \rceil = \lfloor (\frac{\epsilon}{2\log t})^{s+1} \log n \rfloor \geq C_{r,s,\epsilon} \log n \), proving Theorem 1. \( \square \)
3 Density of $K_k$’s in large subsets of the Kneser graph

We are now ready for our main result, a generalization of Theorem 1.3 in [2]:

**Theorem 3** For any fixed $k \in \mathbb{N}$ and $\delta > 0$, if $m \geq 2^{\left(\frac{1}{k+1}+\delta\right)n}$, then any family $\mathcal{G} \subset \mathcal{P}[n]$ of size $|\mathcal{G}| = m$ contains at most

$$\left(\frac{k!}{k^k} + o(1)\right) \binom{m}{k}$$

unordered $k$-tuples $\{A_1, \ldots, A_k\}$ of pairwise disjoint sets, where the $o(1)$ term tends to 0 as $m \to \infty$ for fixed $k, \delta$.

**Proof:**

By increasing $\delta$ if necessary, we may assume $m = 2^{\left(\frac{1}{k+1}+\delta\right)n}$. Consider the subgraph $G$ of the ‘Kneser graph’ on $\mathcal{P}[n]$ induced on the set $\mathcal{G}$, i.e. the graph $G$ with vertex set $\mathcal{G}$ and edge set $\{xy : x \cap y = \emptyset\}$. Let $\epsilon > 0$; we will show that if $n$ is sufficiently large depending on $k, \delta$ and $\epsilon$, the density of $K_k$’s in $G$ is less than $\frac{k!}{k^k} + \epsilon$. Suppose the density of $K_k$’s in $G$ is at least $\frac{k!}{k^k} + \epsilon$; we will obtain a contradiction for $n$ sufficiently large. Let $l = m^f$ (we will choose $f < \frac{\delta}{2(k+1)\delta}$ maximal such that $m^f$ is an integer). By the argument above, there are at least $\frac{\epsilon}{2} \binom{m}{l}$ $l$-sets $U$ such that $G[U]$ has $K_k$-density at least $\frac{k!}{k^k} + \frac{\epsilon}{2}$. Provided $m$ is sufficiently large depending on $k, \delta$ and $\epsilon$, by Theorem 1, each such $G[U]$ contains a copy of $K := K_{k+1}(t)$ where $t \geq C_{k,k,\epsilon/2} \log l = f C'_{k,\epsilon} \log m = C'_{k,\delta,\epsilon} \log m$. Any copy of $K$ is contained in $\binom{m - (k+1)t}{l-1-(k+1)t}$ $l$-sets, so $G$ must contain at least

$$\frac{\epsilon}{2^{\left(m^f - (k+1)^f\right)}} \geq 2^{\left(m/l\right)^{(k+1)f}}$$

copies of $K$.

But we also have the following lemma of Alon and Frankl (Lemma 4.3 in [2]), whose proof we include for completeness:

**Lemma 4** $G$ contains at most $(k+1)2^n(1-\delta t)\binom{m}{t}^{k+1} \frac{1}{(k+1)!}$ copies of $K_{k+1}(t)$.

**Proof:**

The probability that a $t$-subset $\{A_1, \ldots, A_t\}$ chosen uniformly at random from $\mathcal{G}$ has union of size at most $\frac{n}{k+1}$ is at most

$$\sum_{S \subseteq [n] : |S| \leq n/(k+1)} \binom{2|S|}{t}/\binom{m}{t} \leq 2^n (2^{n/(k+1)}/m)^t = 2^n(1-\delta t)$$

Choose at random $k+1$ such $t$-sets; the probability that at least one has union of size at most $n/(k+1)$ is at most

$$(k+1)2^n(1-\delta t)$$
But this condition holds if our \( k + 1 \) \( t \)-sets are the vertex classes of a \( K_{k+1}(t) \) in \( G \). Hence, the number of copies of \( K_{k+1}(t) \) in \( G \) is at most
\[
(k + 1)2^{n(1-\delta t)} \binom{m}{t}^{k+1} \frac{1}{(k+1)!}
\]
as required. \( \square \)

If \( m \) is sufficiently large depending on \( k, \delta \) and \( \epsilon \), we may certainly choose \( t \geq \lceil \frac{4}{\delta} \rceil \), and comparing our two bounds gives
\[
\epsilon \leq 2^{n(1-\delta t)} \binom{m}{k}^{(k+1)t}
\]
Substituting in \( l = m^t \), we get
\[
\epsilon \leq 2^{n(1-\delta t)} m^{f(k+1)t}
\]
Substituting in \( m = 2\left(\frac{k+1+\delta}{k}\right)^n \), we get
\[
\epsilon \leq 2^{n(1-t(\delta-f(1+(k+1)\delta)))} \leq 2^{-n}
\]
since we chose \( f < \frac{\delta}{2(1+(k+1)\delta)} \) and \( t \geq 4/\delta \). This is a contradiction if \( n \) is sufficiently large, proving Theorem 3. \( \square \)

As explained above, our result on \( k \)-generators quickly follows:

**Theorem 5** For fixed \( k \in \mathbb{N} \), any \( k \)-generator \( \mathcal{G} \) of \( \mathbb{P}[n] \) must contain at least \( k2^{n/k}(1-o(1)) \) sets.

*Proof:* Let \( \mathcal{G} \) be a \( k \)-generator of \( \mathbb{P}[n] \), with \( |\mathcal{G}| = m \). As observed in the introduction, the trivial bound gives \( m \geq 2^{n/k} \), so applying Theorem 4 with \( \delta = 1/k(k+1) \), we see that the number of ways of choosing \( k \) pairwise disjoint sets in \( \mathcal{G} \) is at most
\[
\binom{k!}{k^k + o(1)} \binom{m}{k}
\]
The number of ways of choosing less than \( k \) pairwise disjoint sets is, very crudely, at most \( \sum_{i=0}^{k-1} \binom{m}{i} = \Theta(1/m) \binom{m}{k} \); since every subset of \( [n] \) is a disjoint union of at most \( k \) sets in \( \mathcal{G} \), we obtain
\[
2^n \leq \left( \frac{k!}{k^k + o(1) + \Theta(1/m)} \right) \binom{m}{k} = \left( \frac{m}{k} \right)^k (1 + o(1))
\]
(where the \( o(1) \) term tends to 0 as \( m \to \infty \)), and therefore
\[
m \geq k2^{n/k}(1-o(1))
\]
(where the \( o(1) \) term tends to 0 as \( n \to \infty \)). \( \square \)

**Note:** The author wishes to thank Peter Keevash for bringing to his attention the result of Erdős in [3], after reading a previous draft of this paper in which a weaker, asymptotic version of Erdős’ result was proved.
References

[1] Frein, Y., Lévêque, B., Sebő, A., Generating All Sets With Bounded Unions, *Combinatorics, Probability and Computing* 17 (2008) pp. 641-660

[2] Alon, N., Frankl, P., The Maximum Number of Disjoint Pairs in a Family of Subsets, *Graphs and Combinatorics* 1 (1985), pp. 13-21

[3] Erdős, P., On the number of complete subgraphs contained in certain graphs, *Publ. Math. Inst. Hung. Acad. Sci., Ser. A* 7 (1962), pp. 459-464

[4] Nikiforov, V., Graphs with many $r$-cliques have large complete $r$-partite subgraphs, *Bulletin of the London Mathematical Society* Volume 40, Issue 1 (2008) pp. 23-25

[5] Furedi, Z., Katona, G.O.H., 2-bases of quadruples, *Combinatorics, Probability and Computing* 15 (2006) pp. 131-141