Comparable pairs in families of sets

Noga Alon, Shagnik Das, Roman Glebov, Benny Sudakov

Definitions and results

Let $A, B \subset [n] = \{1, 2, \ldots, n\}$ be two different sets. We say that sets A, B *are comparable* if either $A \subset B$ or $B \subset A$. Let $\mathcal{F}$ be a family od subsets of $[n]$, i.e. $\mathcal{F} \subseteq 2^{[n]}$. Denote by $c(\mathcal{F})$ the number of comparable pairs in $\mathcal{F}$ and define $c(n, m)$ to be the maximum number of comparable pairs over all families of m subsets of $[n]$.

Clearly $c(n, m) \leq \binom{m}{2}$ and equality holds only for $m \leq n + 1$. Again, little counting shows $c(n, 2^n) = 3^n - 2^n \sim 2^{1.58n}$.

Daykin and Frankl showed that there are "large" families where almost all pairs are comparable. More precisely, they proved that $c(n, m) = (1 - o(1))\binom{m}{2}$ if and only if $m = 2^{o(n)}$. Their lower bound comes from the construction called *the tower of cubes*, which generalizes linearly ordered sets. A *subcube* of a hypercube $2^{[n]}$ is, for some sets $F_1 \subset F_2$, a family of sets $\mathcal{F} = \{F \subset [n] : F_1 \subseteq F \subseteq F_2\}$. We call the number $|F_2| - |F_1|$ the *dimension* of $\mathcal{F}$. For simplicity suppose that $k \mid n$ and $l = n/k$. Let $X_i = [il]$ for $i \in [k]$, and consider family

$$\mathcal{F}_i = \{F \subset [n] : X_{i-1} \subseteq F \subseteq X_i\} \quad \text{and} \quad \mathcal{F} = \cup_{i=1}^{k} \mathcal{F}_i.$$

Then $|\mathcal{F}| = k2^{n/k} - k + 1$; moreover, any two sets from different subcubes are comparable, hence $c(\mathcal{F}) \geq (1 - 1/k)\binom{m}{2}$. When $k = \omega(1)$, then $c(\mathcal{F}) = (1 - o(1))\binom{m}{2}$ as in the theorem of Daykin and Frankl.

Alon and Frankl proved that the towers of cubes are asymptotically optimal even when k is constant.

Theorem 1 For every positive integer k there exists a positive $\beta = \beta(k)$ such that if $m = 2^{(1/(k+1)+\delta)n}$ for $\delta > 0$, then

$$c(n, m) < \left(1 - \frac{1}{k}\right)\binom{m}{2} + O\left(m^{2-\beta\delta^{k+1}}\right).$$

Case $k = 1$ is of particular interest. Alon and Frankl proved that a family $\mathcal{F}$ of size $m = 2^{(1/2+\delta)n}$ must have $c(n, m) < 4m^{2-\delta^2/2} = o(m^2)$, thus proving a conjecture of Daykin and Erdős. Erdős made a finer conjecture, asking whether $m = \omega(2^{n/2})$ implies $c(n, m) = o(m^2)$. Alon and Frankl disproved this conjecture, exhibiting for any $d \geq 1$ a family $\mathcal{F}$ of size $\Omega(n^d 2^{n/2})$ with $c(\mathcal{F}) \geq 2^{-2d-1}\binom{m}{2}$, and they conjectured that this construction is essentially the best possible.

Conjecture 1 If $m = n^{\omega(1)} 2^{n/2}$ then $c(n, m) = o(m^2)$.

Proving a conjecture of Alon and Frankl

We prove a slightly more general result. Let $\mathcal{A}$ and $\mathcal{B}$ be two families of subsets of $[n]$, we write $c(\mathcal{A}, \mathcal{B})$ for the number of pairs $(A, B) \in \mathcal{A} \times \mathcal{B}$ with $A \subset B$.

Theorem 2 Let $\mathcal{A}$ and $\mathcal{B}$ be two set families over $[n]$ with $|\mathcal{A}||\mathcal{B}| = n^d 2^n$. Then $c(\mathcal{A}, \mathcal{B}) \leq 2^{-d/300}|\mathcal{A}||\mathcal{B}|$.

For the proof we need following lemmata.

Lemma 1 For $n \geq 2$ and $d \in (0,1]$, let $\mathcal{A}$ and $\mathcal{B}$ be two families of subsets of $[n]$ satisfying $c(\mathcal{A},\mathcal{B}) \geq 2^{-d/300}|\mathcal{A}||\mathcal{B}|$. Then $|\mathcal{A}||\mathcal{B}| < n^d 2^n$.

Lemma 2 For $p, q \in [0,1]$ and $0 < \alpha < 1/300$ we have

$$(p(1-q))^{1-\alpha} + (pq)^{1-\alpha} + ((1-p)q)^{1-\alpha} \leq \left(2 + 2^{1-\frac{1}{300\alpha}}\right)^{\alpha}.$$

Lemma 3 Let $\mathcal{F}$ be a family of subsets of $[n]$ and let p_i denote the fraction of sets in $\mathcal{F}$ that contain i. Then

$$|\mathcal{F}| \leq 2^{\sum_{i=1}^{n} H(p_i)},$$

where $H(p) = -p\log p - (1-p)\log(1-p)$ is the binary entropy.

Dense families

Here we describe the nature of large families $\mathcal{F}$ maximizing $c(\mathcal{F})$. Since the sets $\emptyset$ and $[n]$ are contained in all possible comparable pairs, one can guess that families $\mathcal{F}$ maximizing $c(\mathcal{F})$ are concentrated close to the sets $\emptyset$ and $[n]$. We show that it is the case for $|\mathcal{F}| \geq 2^{0.92n}$.

Let $0 \leq k \leq n/2$, we define

$$\mathcal{H}_k = \{F \subset [n] : |F| \leq k\} \cup \{F \subset [n] : |F| \geq n-k\},$$
$$M_k = |\mathcal{H}_k| = 2\sum_{i=0}^{k} \binom{n}{i}.$$

Theorem 3 If $M_{k-1} \leq m \leq M_k$ for some k with $n/3 + \sqrt{2n\ln 2} \leq k \leq n/2$, then every family $\mathcal{F}$ of m sets over $[n]$ maximising the number of comparable pairs satisfies $\mathcal{H}_{k-1} \subseteq \mathcal{F} \subseteq \mathcal{H}_k$.

The entropy bound $2^{0.92n}$ comes from the estimate $\sum_{i\leq pn}\binom{n}{i} \leq 2^{H(p)n}$. The proof is based on switching sets between $\mathcal{F}$ and $2^{[n]} \setminus \mathcal{F}$.

Lemma 4 If $\lambda \geq 0$ and $\mathcal{A} = \{A \subseteq [r] : |A| \geq r/2 + \lambda\sqrt{r}\}$, then $|\mathcal{A}| \leq e^{-2\lambda^2}2^r$.

Lemma 5 Given integers n an s satisfying $n/3 < s \leq n/2$, the quantity

$$2^{n-t} + \sum_{j=0}^{s}\binom{t}{j}$$

is minimised over $n - s \leq t \leq n$ when $t = n - s$.

The above theorem is enough to determine $c(n,m)$ asymptotically, but we wish to determine the exact value of $c(n,m)$. That can be done for small values of $m' = m - M_{k-1}$ by following proposition, but remains open for arbitrary m'. However, the exact result has some subtle complexities, since it contains as a special case the famous Kruskal-Katona Theorem.

Proposition 1 Suppose $n/3+\sqrt{2n\ln 2} \leq k \leq n/2$ and $m' = m - M_{k-1} \leq 2\binom{n/2}{k}$. Then $c(n,m) = c(\mathcal{H}_{k-1} \cup \mathcal{A} \cup \mathcal{B})$, where $\mathcal{A}$ is a set of $\lfloor m'/2 \rfloor$ k-subsets of $[n/2]$ and $\mathcal{B}$ is a set of $\lceil m'/2 \rceil$ $(n-k)$-sets containing $[n/2]$.