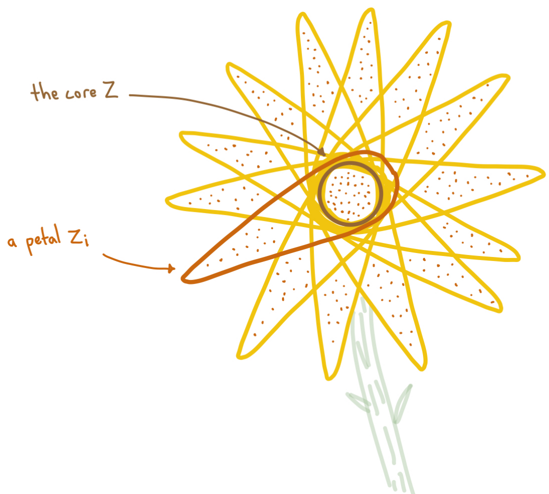
Sunflower Lemma

Styopa Zharkov

Definition: A collection of sets $\{Z_1, \ldots, Z_p\}$ is called a *sunflower* if there exists a set Z (which we call the *core*) such that for any $i \neq j$, we have $Z_i \cap Z_j = Z$. The sets Z_i are called the *petals* of the sunflower[1]. Equivalently, every element either belongs to none, one, or all of the sets in the collection. ◁

Note that a collection where all sets are disjoint from each other is also a sunflower with an empty core.

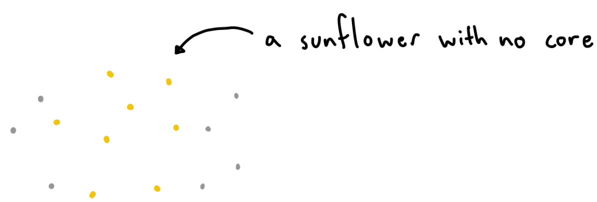


Lemma (Sunflower): Let $\mathcal{T}$ be a collection of non-empty sets each of size at most ℓ. If $\mathcal{T}$ contains more than $\ell!(p-1)^\ell$ sets, then it contains a sunflower with p petals. ◁

Intuitively, this means that large collections of small sets have sunflowers.

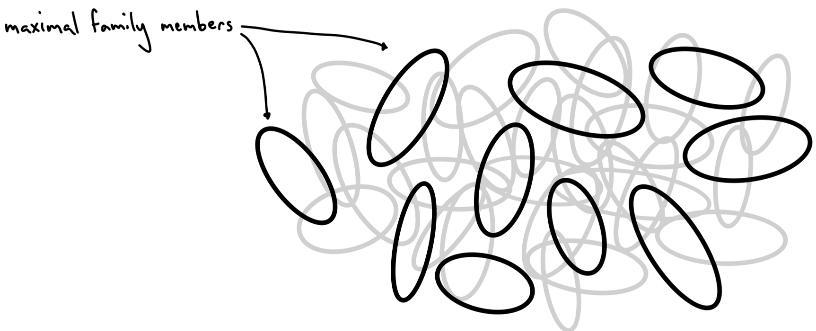
Proof: We will prove this lemma by induction on ℓ.

For the base case, if $\ell = 1$, then all sets must contain only one element. If $|\mathcal{T}| > \ell!(p-1)^\ell = p-1$, then we can choose any p sets in $\mathcal{T}$. This is a sunflower with the empty set as its core.



[1] Here, we recklessly call the sets in the sunflower petals, even though real sunflower cores aren't considered to be part of their petals.

Now, let's show the inductive step. Let $\ell \geq 2$. Suppose that we know that the lemma holds for $\ell - 1$. Let's take a maximal family[2] of pairwise disjoint sets in $\mathcal{T}$. Let the sets in the family be called $X_1, \ldots, X_t$ and let $X = X_1 \cup \cdots \cup X_t$.



If $t \geq p$, then we are done because any p sets from our family form a sunflower with an empty core.

If $t < p$, then we can see that

$$|X| = \sum_{i=0}^{t} |X_i| \leq \sum_{i=0}^{t} \ell = \ell(p-1).$$

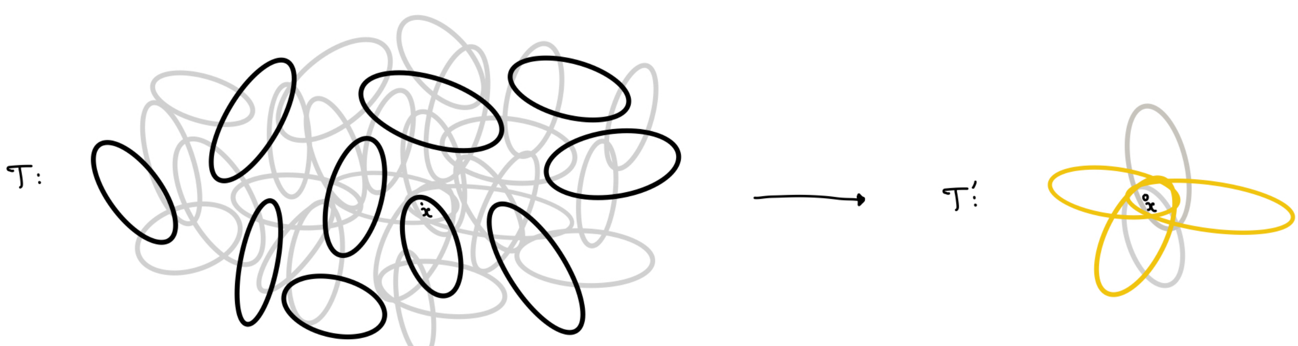
Since our family is maximal, every set in $\mathcal{T}$ must intersect with X (otherwise we would be able to add it to the collection). There are more than $\ell!(p-1)^\ell$ sets in $\mathcal{T}$ and only $\ell(p-1)$ points in X, by the pigeonhole principle, there exists some point $x \in X$ that is contained in at least

$$\frac{|\mathcal{T}|}{|X|} \geq \frac{\ell!(p-1)^\ell}{\ell(p-1)} = (\ell-1)!(p-1)^{\ell-1}$$

of the sets in $\mathcal{T}$. If we remove x from all of the sets containing it, then those sets form a collection on which we can use our inductive hypothesis. More formally, if we take the set

$$\mathcal{T}' = \{S \setminus \{x\} \colon S \in \mathcal{T}, x \in S\},$$

then all sets in $\mathcal{T}'$ have at most $\ell - 1$ elements, so by the inductive hypothesis, there is a sunflower with p petals in $\mathcal{T}'$.



If we add x back to all of the sets in this sunflower, then we still have a sunflower with p petals, but now all sets are in $\mathcal{T}$. This is what we wanted to prove existed, so we are done. □

[2] Here, maximal means that it's impossible to add another set to our family while maintaining the property that all sets are pairwise disjoint.