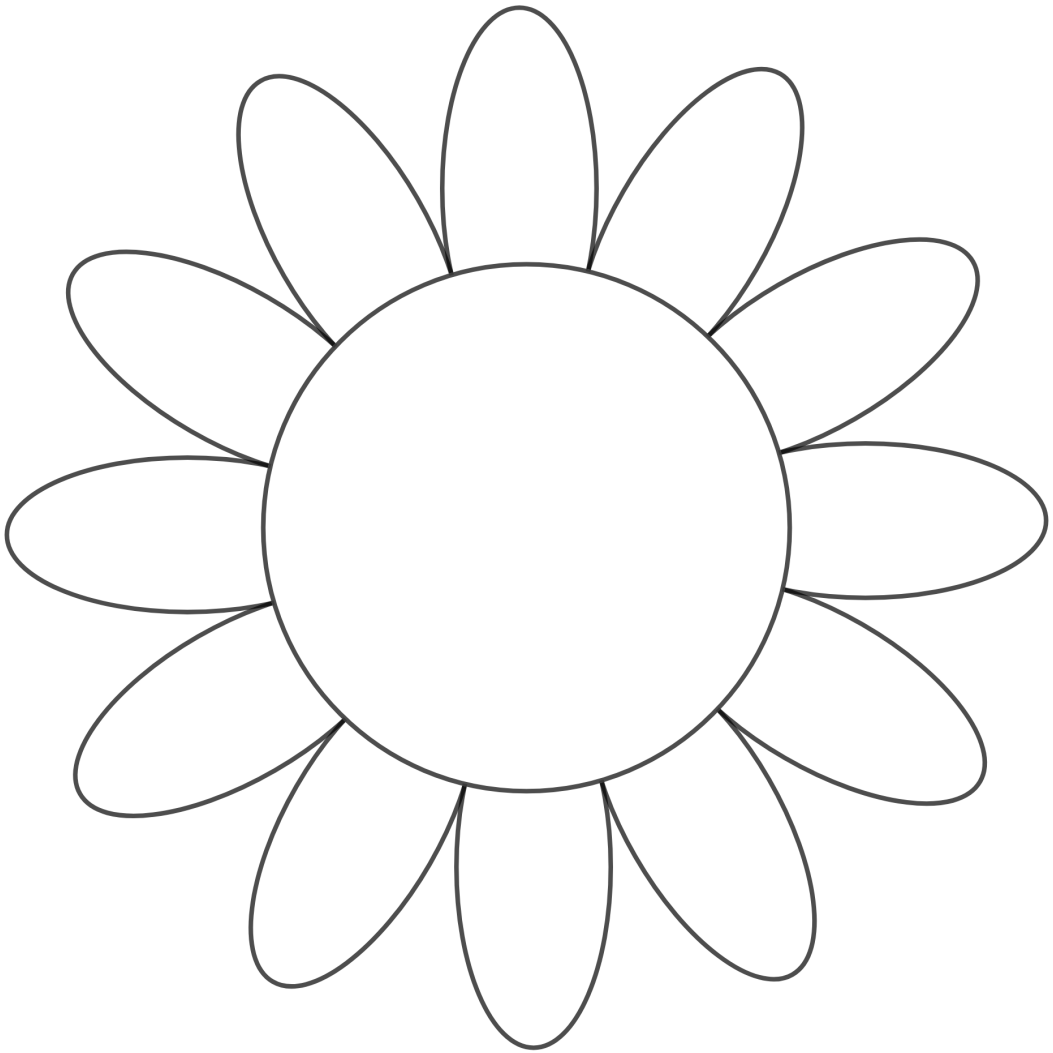
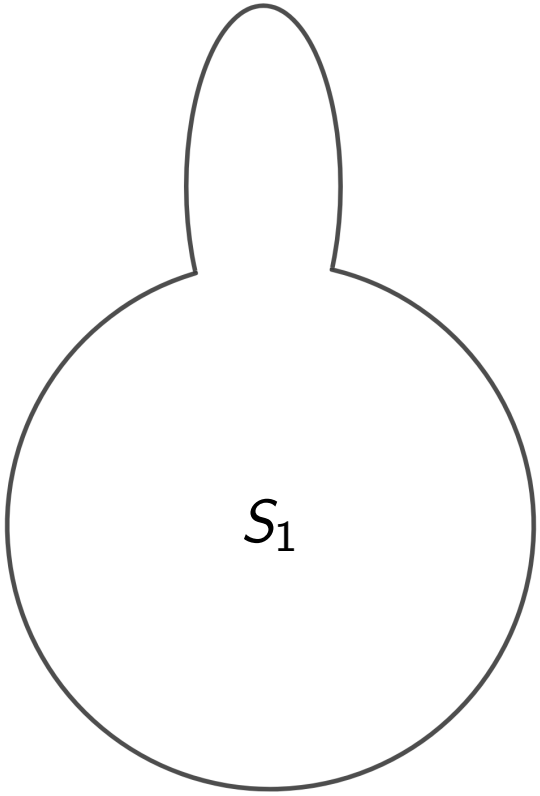
Sunflowers, daisies and local codes

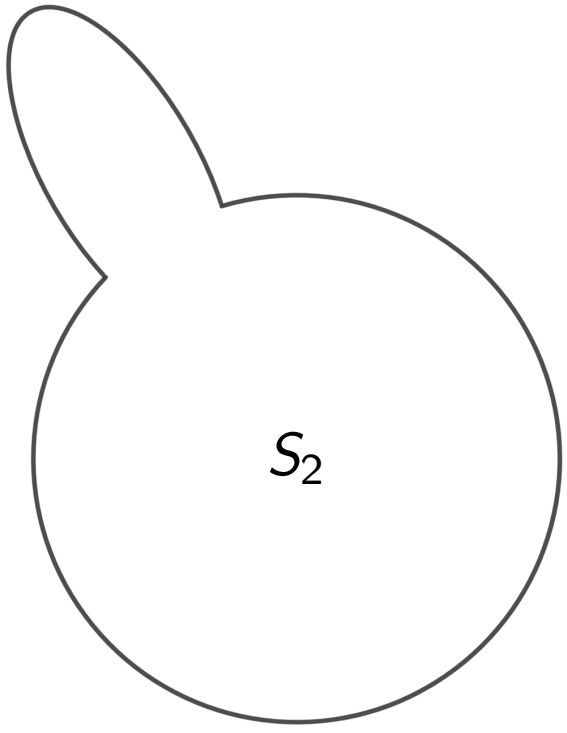
Marcel de Sena
(joint with Tom Gur and Oded Lachish)

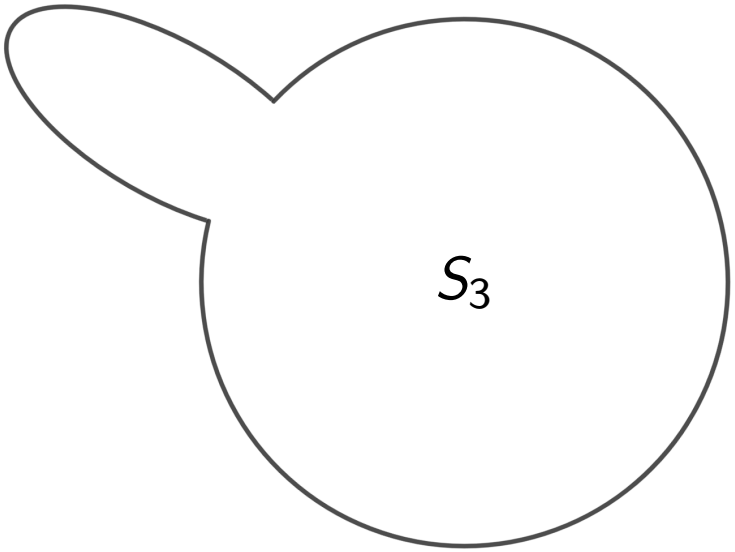
By KMJ, CC BY-SA 3.0. URL: https://commons.wikimedia.org/w/index.php?curid=3301347

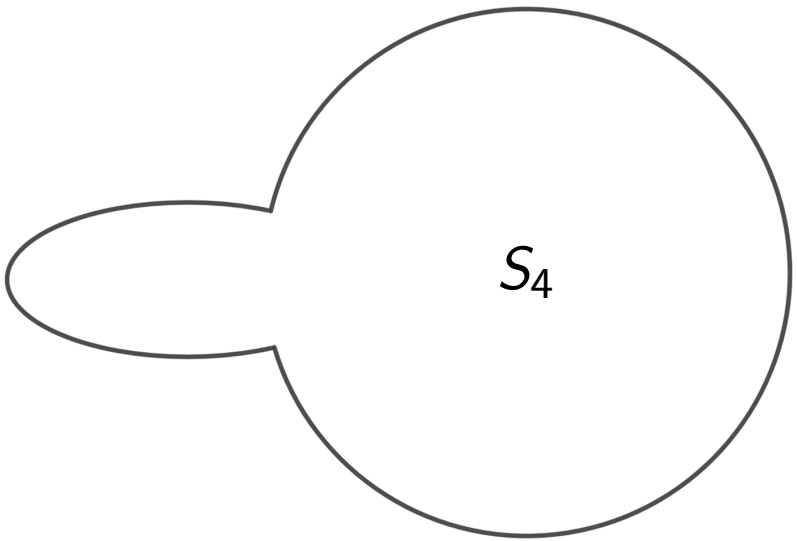


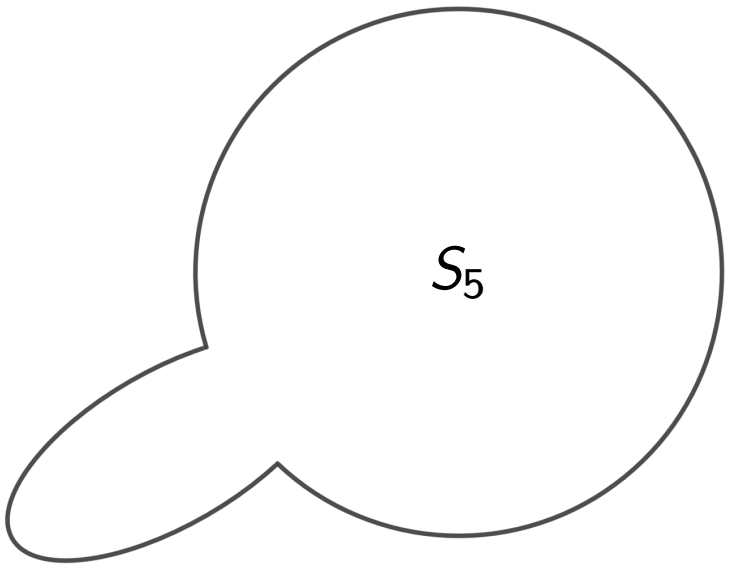
$$\mathcal{S} = \{S_1, S_2, \ldots, S_{12}\}$$

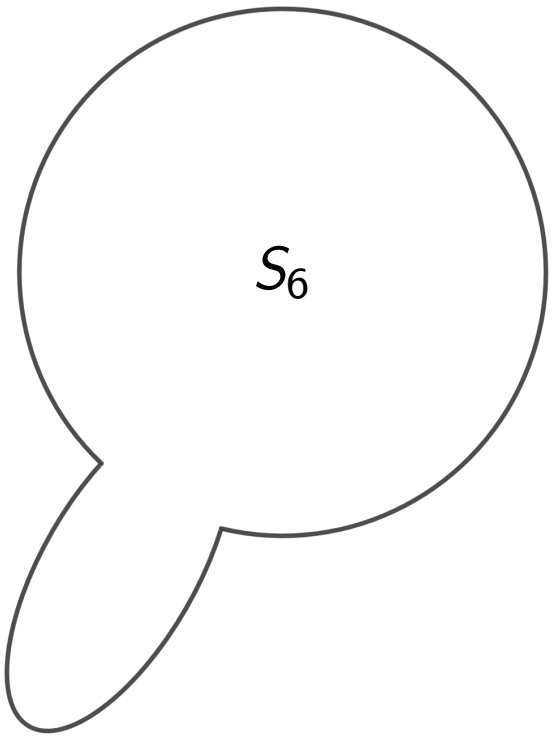

S_1

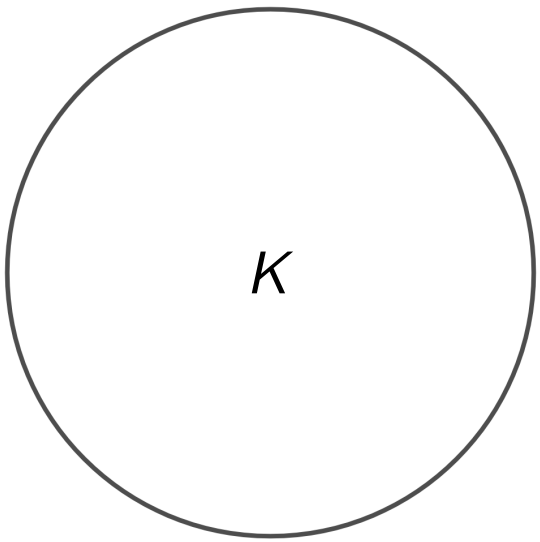

S_2

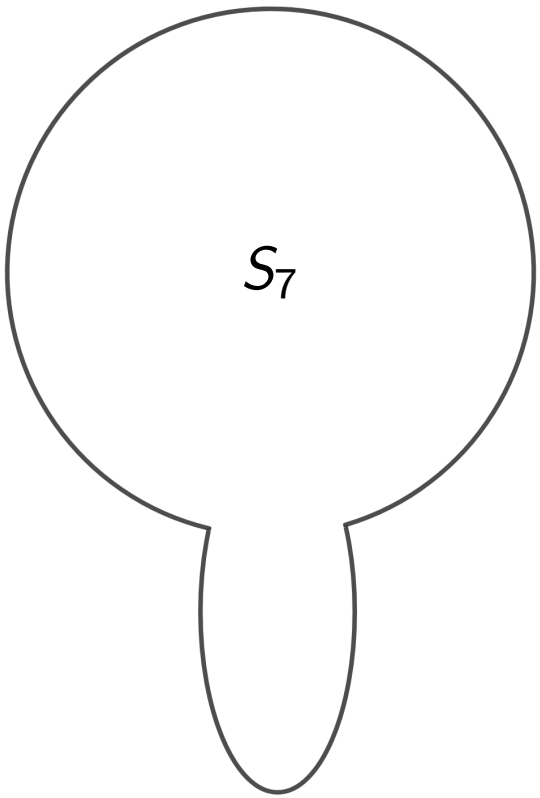

S_3


S_4

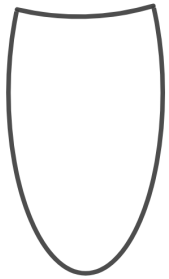

S_5


S_6


K


S_7

$$P_7 = S_7 \setminus K_7$$



Definition

A *sunflower* is a collection $\mathcal{S}$ of q-sets such that $S \cap S' = \cap_{T \in \mathcal{S}} T = K$ for any distinct $S, S' \in \mathcal{S}$. The set K is called the *kernel*, and each $P = S \setminus K$ is a *petal*.

Sunflower lemma [ER60]

If $|\mathcal{S}| \geq q!(s-1)^q = \Theta(sq)^q$, then $\mathcal{S}$ contains a sunflower of size s.

Sunflower lemma [ER60]

If $|\mathcal{S}| \geq q!(s-1)^q = \Theta(sq)^q$, then $\mathcal{S}$ contains a sunflower of size s.

Sunflower conjecture [ER60]

For some $c : \mathbb{N} \to \mathbb{N}$, if $|\mathcal{S}| \geq c(s)^q$, then $\mathcal{S}$ contains a sunflower of size s.

Sunflower lemma [ER60]

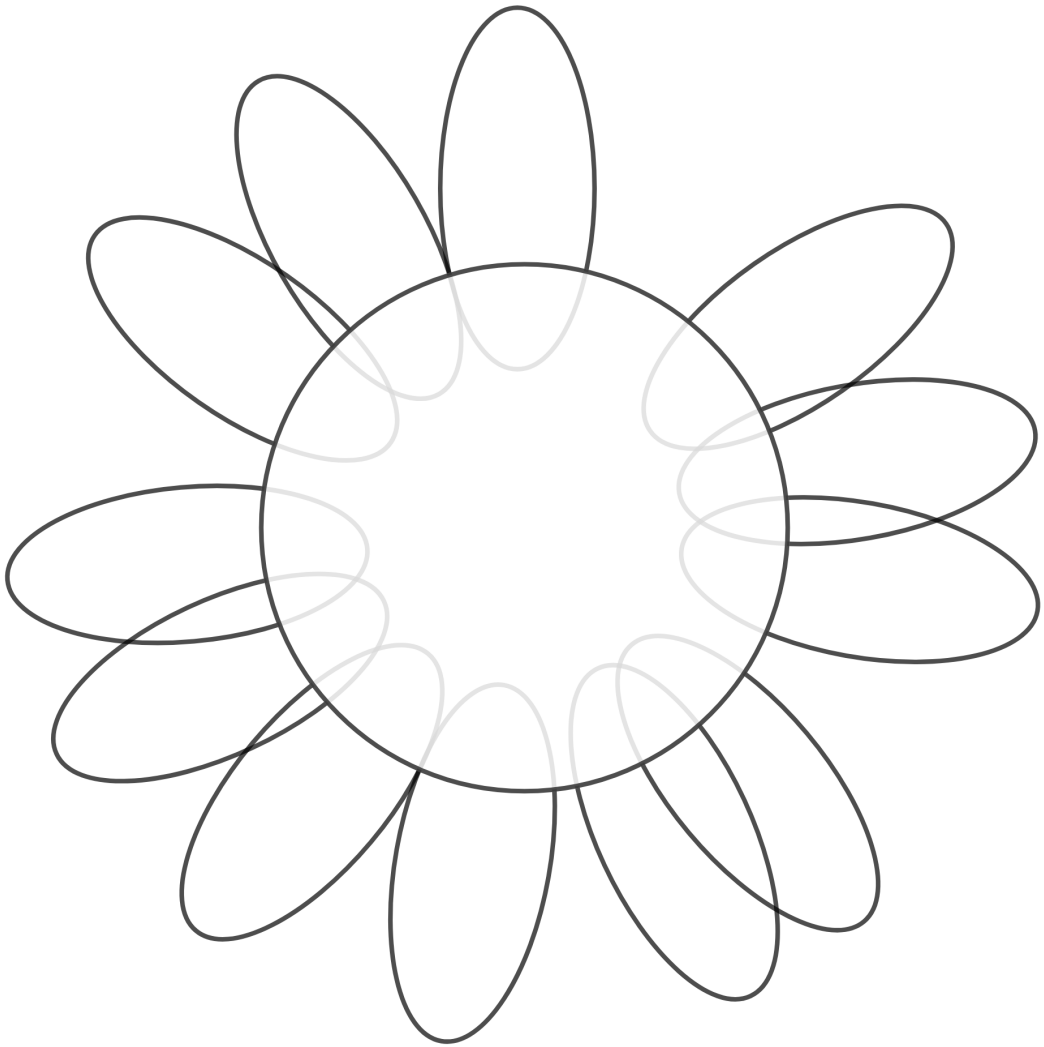
If $|\mathcal{S}| \geq q!(s-1)^q = \Theta(sq)^q$, then $\mathcal{S}$ contains a sunflower of size s.

Sunflower conjecture [ER60]

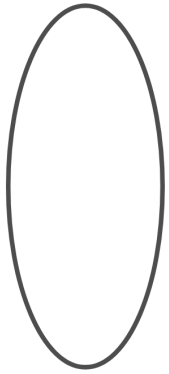
For some $c : \mathbb{N} \to \mathbb{N}$, if $|\mathcal{S}| \geq c(s)^q$, then $\mathcal{S}$ contains a sunflower of size s.

Theorem [ALWZ19]

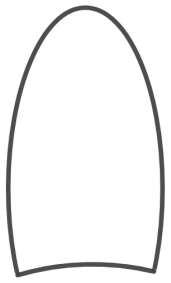
If $|\mathcal{S}| = \Omega(s \log q)^q$, then $\mathcal{S}$ contains a sunflower of size s.



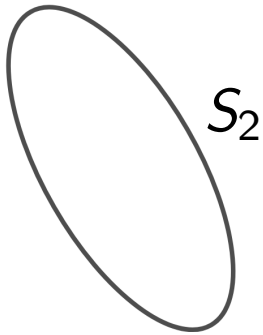
$$\mathcal{D} = \{S_1, S_2, \ldots, S_{12}\}$$

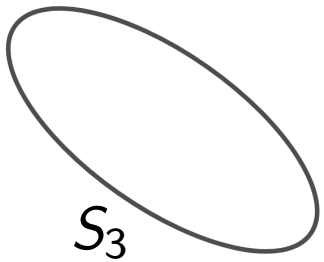


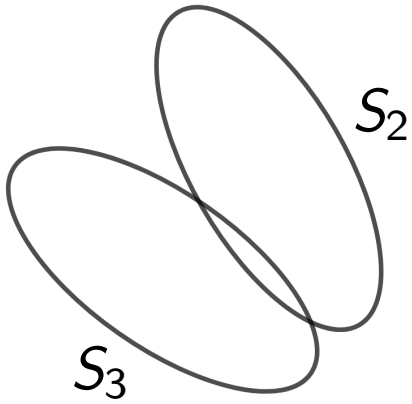
S_1



$$P_1 = S_1 \setminus K$$


S_2

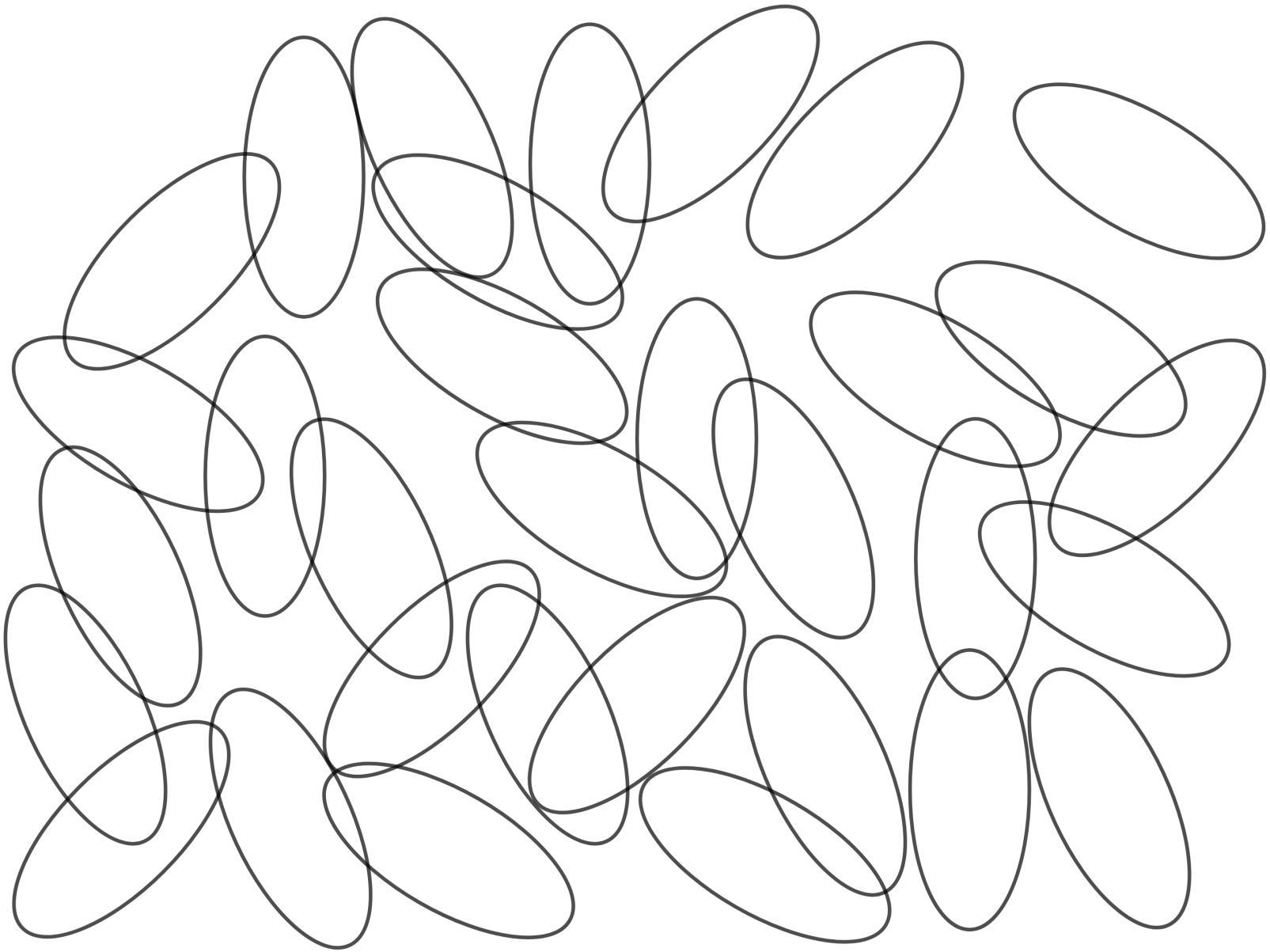

S_3


S_2
S_3

Definition

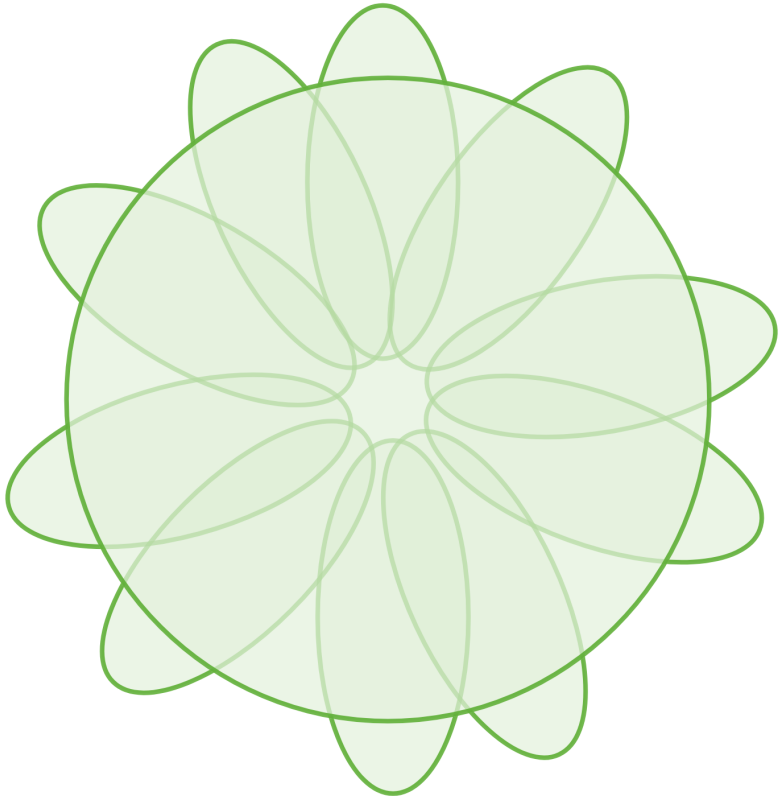
A *t-daisy* with *kernel* K is a collection $\mathcal{D}$ of q-sets such that each $i \notin K$ is contained in at most t members of $\mathcal{D}$.

Daisy partition lemma

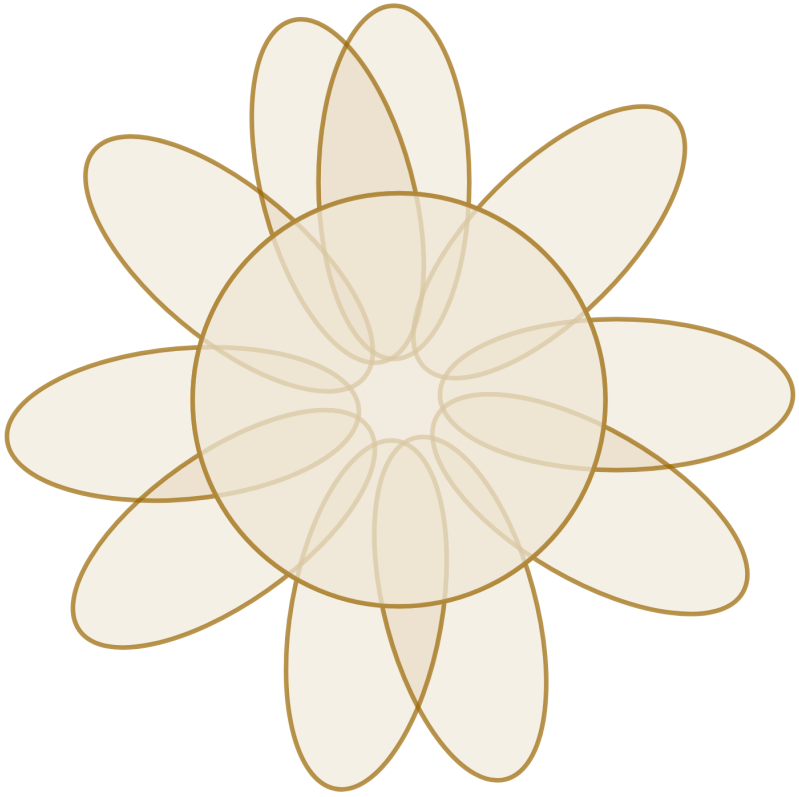


Daisy partition lemma

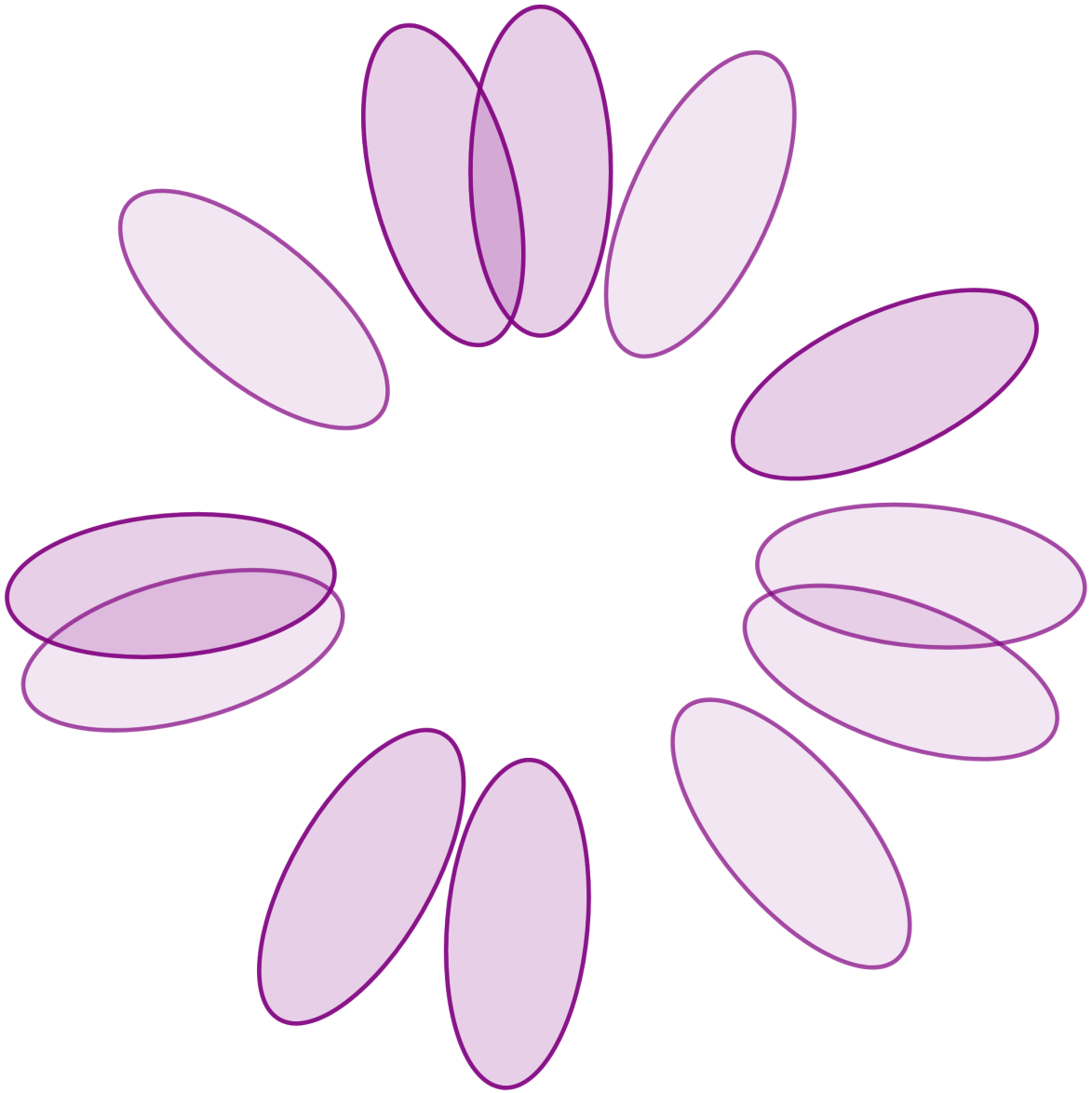
Daisy partition lemma



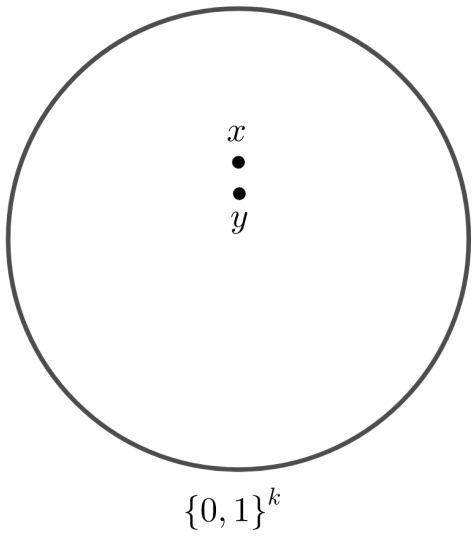
Daisy partition lemma



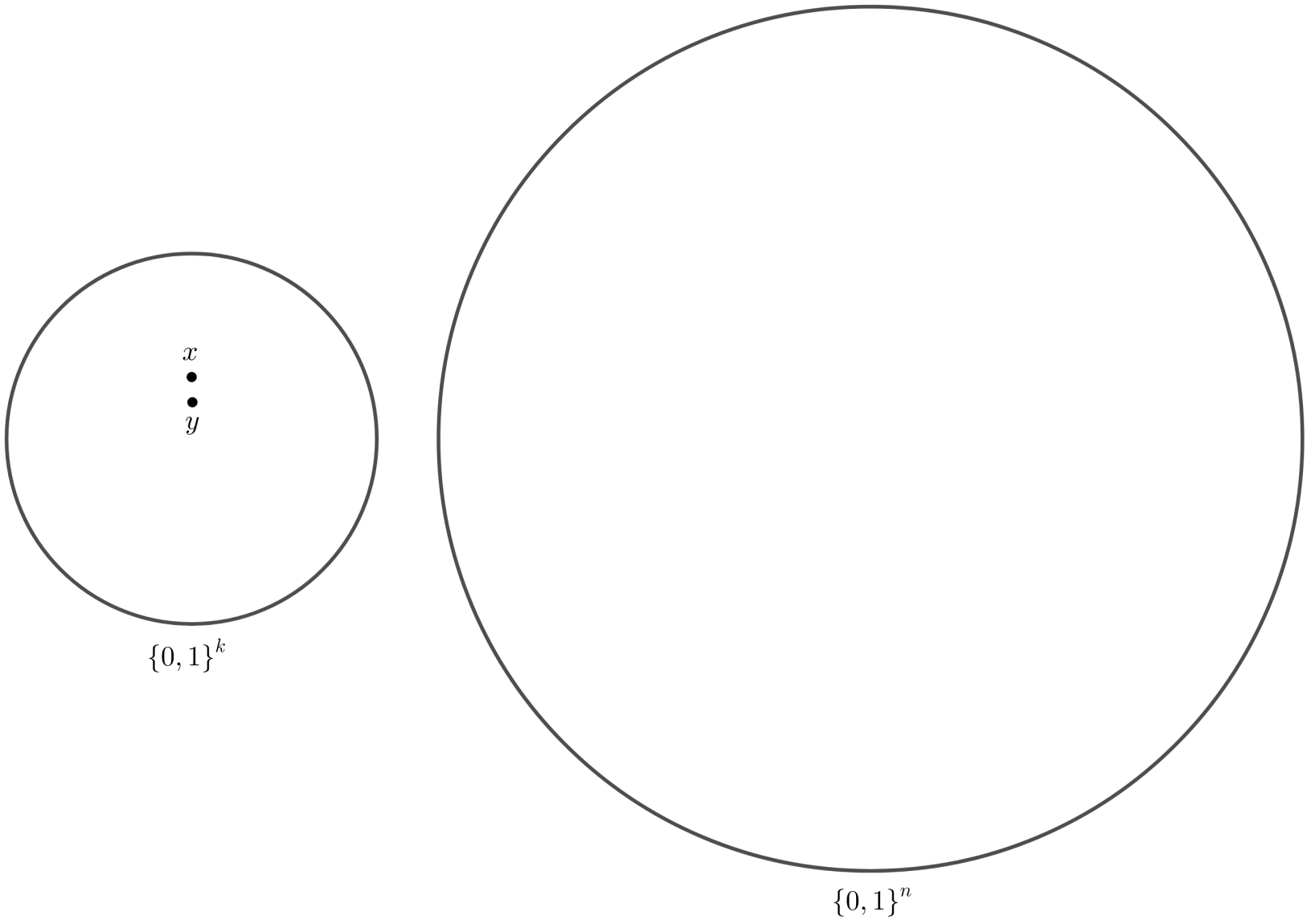
Daisy partition lemma



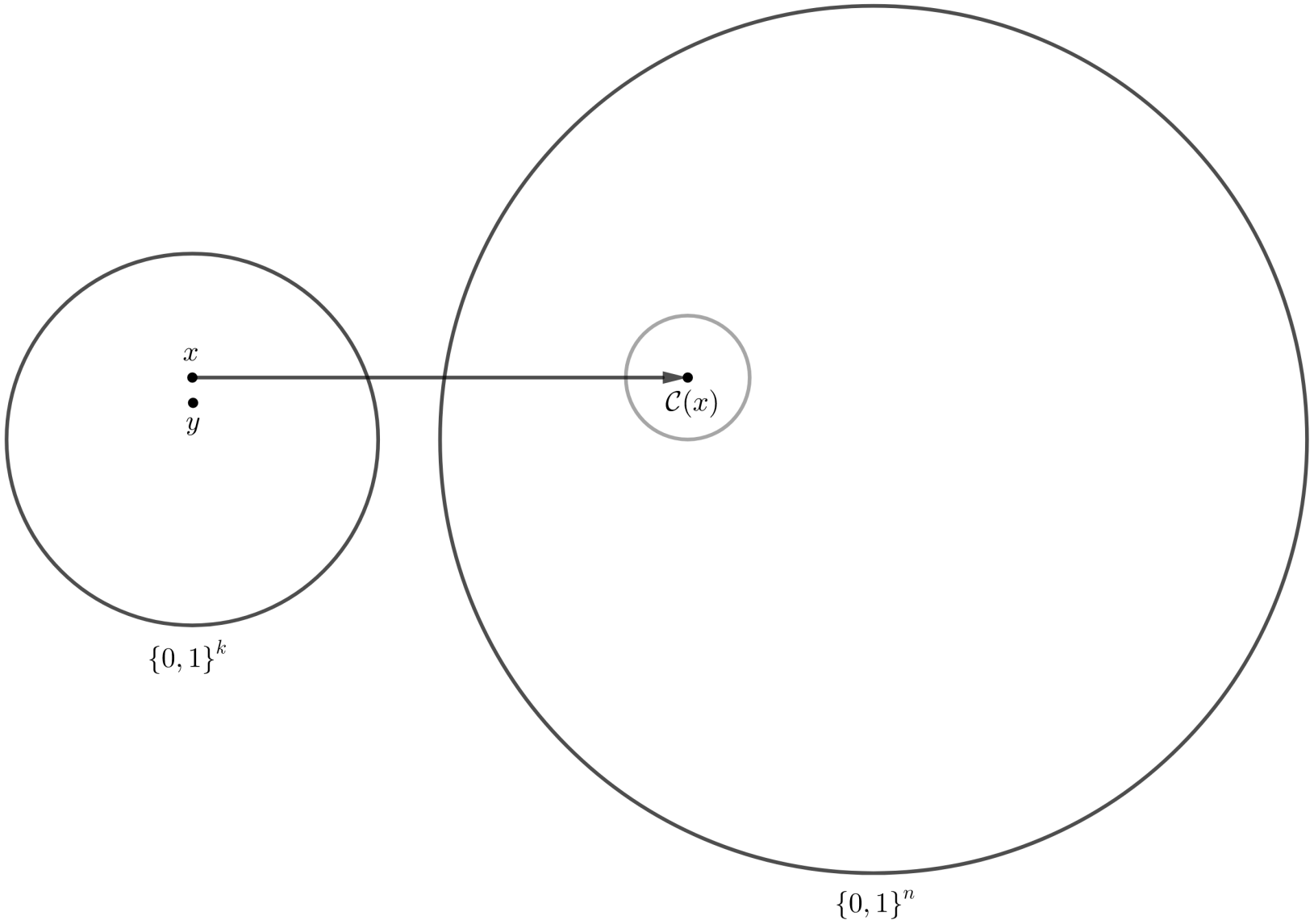
Error-correcting codes



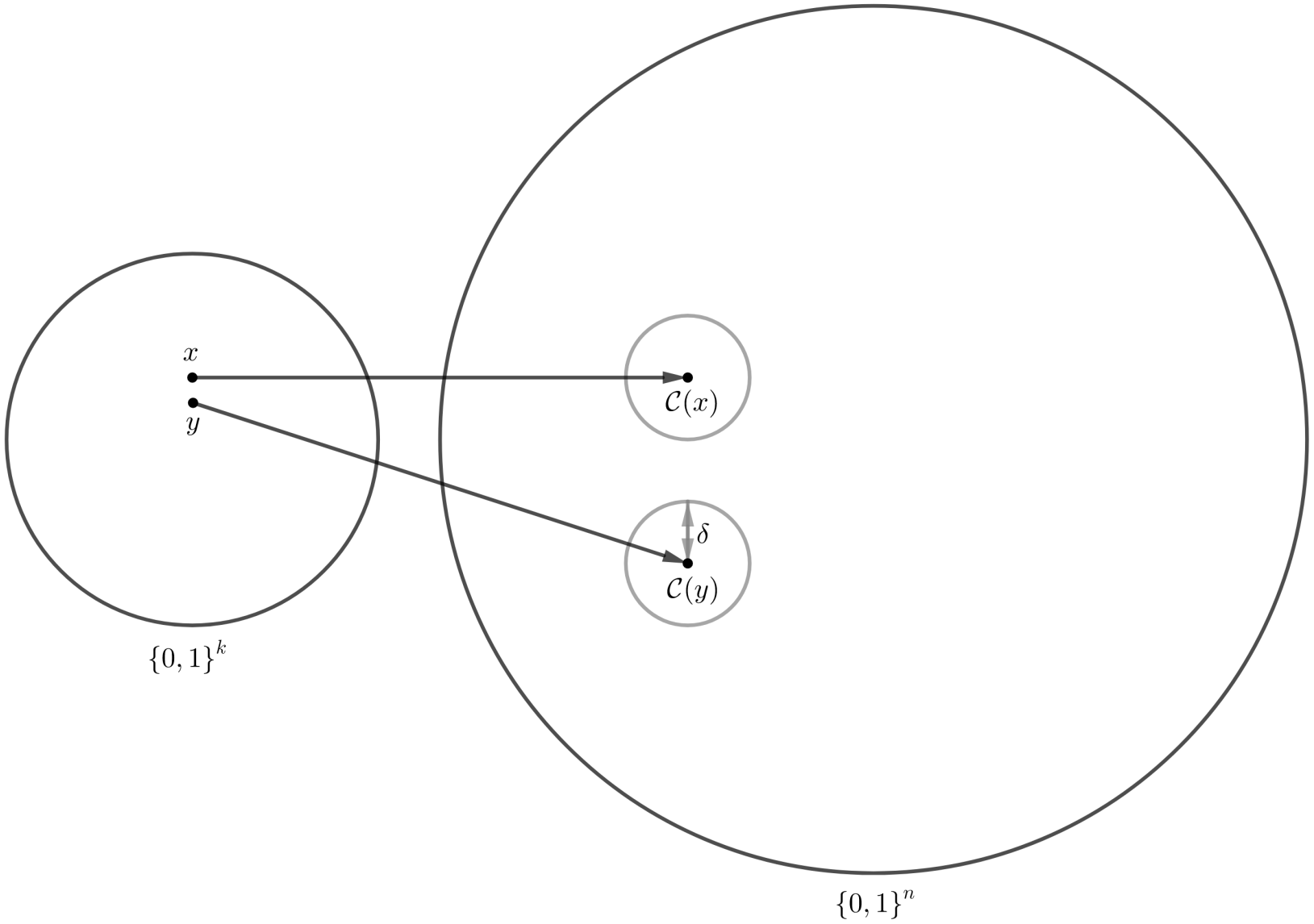
Error-correcting codes



Error-correcting codes



Error-correcting codes



Error-correcting codes

Definition

A binary *error-correcting code* is an injective function $\mathcal{C} : \{0,1\}^k \to \{0,1\}^n$ where the preimage (message) is recoverable after significant corruption of the image (codeword).

Error-correcting codes

Definition

A binary *error-correcting code* is an injective function $\mathcal{C} : \{0,1\}^k \to \{0,1\}^n$ where the preimage (message) is recoverable after significant corruption of the image (codeword).

- Δ: relative distance (recoverable from Δn corruptions)
- k: message length
- n: blocklength

Error-correcting codes

Codes with large distance are resilient to corruption; codes with large *rate* k/n have little redundancy. **Goal: find codes with high rate and distance.**

Error-correcting codes

Codes with large distance are resilient to corruption; codes with large *rate* k/n have little redundancy. **Goal: find codes with high rate and distance.**

Singleton bound

For any binary code $\mathcal{C}$,

$$\frac{k}{n} + \Delta \leq 1 - \frac{1}{n} < 1.$$

Locally decodable codes

Definition

$\mathcal{C}$ is a *locally decodable code* (LDC) if one need only look at a small number of coordinates of $w \approx \mathcal{C}(x)$ to decode x_i.

Locally decodable codes

Definition

There exists a (randomised) algorithm D with *decoding radius* δ such that, if w is δ-close to $\mathcal{C}(x)$, then, $\forall i$,

$$\mathbb{P}[D^w(i) = x_i] \geq 2/3$$

and D makes $q = o(n)$ queries to w.

Locally decodable codes

Definition

There exists a (randomised) algorithm D with *decoding radius* δ such that, if w is δ-close to $\mathcal{C}(x)$, then, $\forall i$,

$$\mathbb{P}[D^w(i) = x_i] \geq 2/3$$

and D makes $q = O(1)$ queries to w.

Relaxed locally decodable codes

Definition

$\mathcal{C}$ is a *relaxed locally decodable code* (RLDC) if $\mathcal{C}$ is (almost) locally decodable but D can sometimes fail and return $\perp$.

Question

Is there some $\Delta > 0$ for which there exist codes of distance Δ and blocklength $n = O(k)$?

Question

Is there some $\Delta > 0$ for which there exist codes of distance Δ and blocklength $n = O(k)$?

Yes!

Question

Is there some $\Delta > 0$ for which there exist LDCs of distance Δ and blocklength $n = O(k)$?

Question

Is there some $\Delta > 0$ for which there exist RLDCs of distance Δ and blocklength $n = O(k)$?

Question

Is there some $\Delta > 0$ for which there exist RLDCs of distance Δ and blocklength $n = O(k)$?

No!

Theorem [GL19]

Any *one-sided* RLDC $\mathcal{C}$ with message length k and blocklength n satisfies

$$n = k^{1+\Omega\left(\frac{1}{2^q}\right)}.$$

Theorem [GL19]

Any *one-sided* RLDC $\mathcal{C}$ with message length k and blocklength n satisfies

$$n = k^{1+\Omega\left(\frac{1}{2^q}\right)}.$$

Theorem

Any *two-sided* RLDC $\mathcal{C}$ with message length k and blocklength n satisfies

$$n = k^{1+\Omega\left(\frac{1}{q^2}\right)}.$$

Overview

1. Local decoder $D' \to$ "analysable" decoder D
2. From D, obtain *global* decoder G – using daisies!
3. G decodes all k bits *from a valid codeword* with high probability and $\approx n^{1-\alpha}$ queries: information theoretically,

$$n^{1-\alpha} \approx k \implies n \approx k^{1+\beta}.$$

Overview

1. Local decoder $D' \to$ "analysable" decoder D
2. From D, obtain *global* decoder G – using daisies!
3. G decodes all k bits *from a valid codeword* with high probability and $\approx n^{1-\alpha}$ queries: information theoretically,

$$n^{1-\alpha} \approx k \implies n \approx k^{1+\beta}.$$

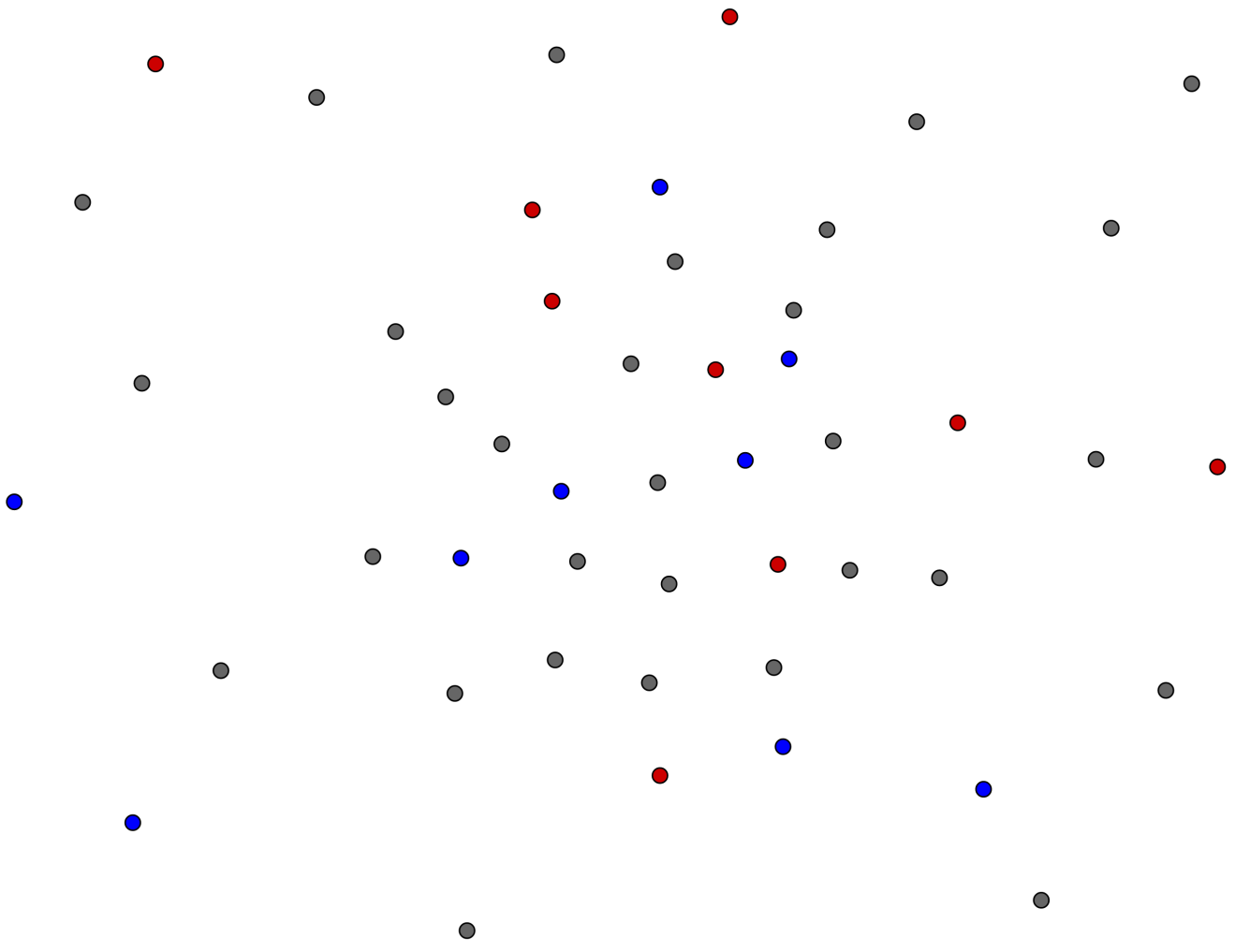
Overview

1. Local decoder $D' \rightarrow$ "analysable" decoder D
2. From D, obtain *global* decoder G – using daisies!
3. G decodes all k bits *from a valid codeword* with high probability and $\approx n^{1-\alpha}$ queries: information theoretically,

$$n^{1-\alpha} \approx k \implies n \approx k^{1+\beta}.$$

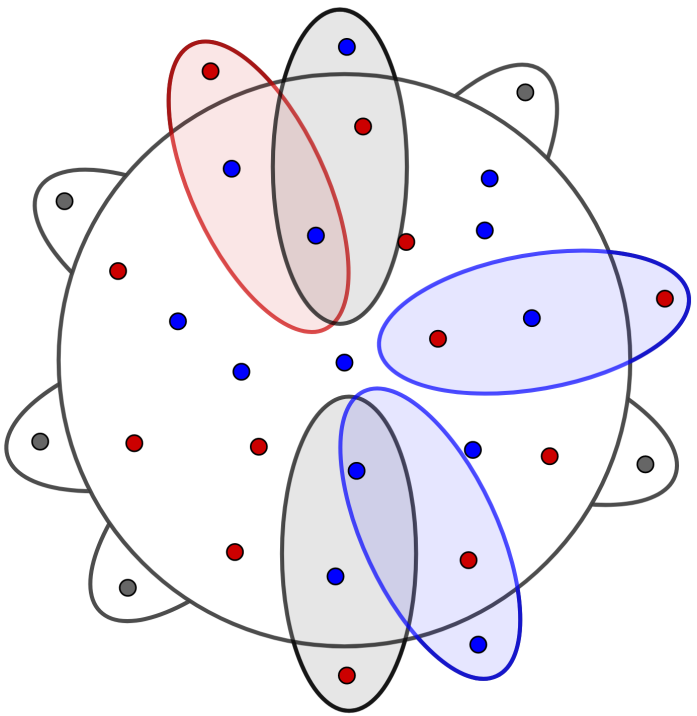
Global decoder G: construction

Binomial sampling with $p = n^{-\frac{1}{2q^2}}$



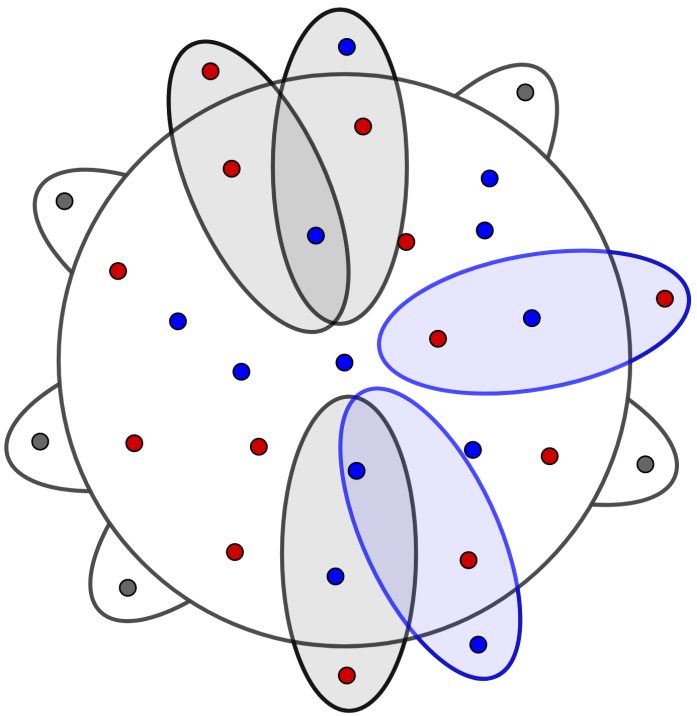
Global decoder G: construction

$\mathcal{D}_1$, assignment 1



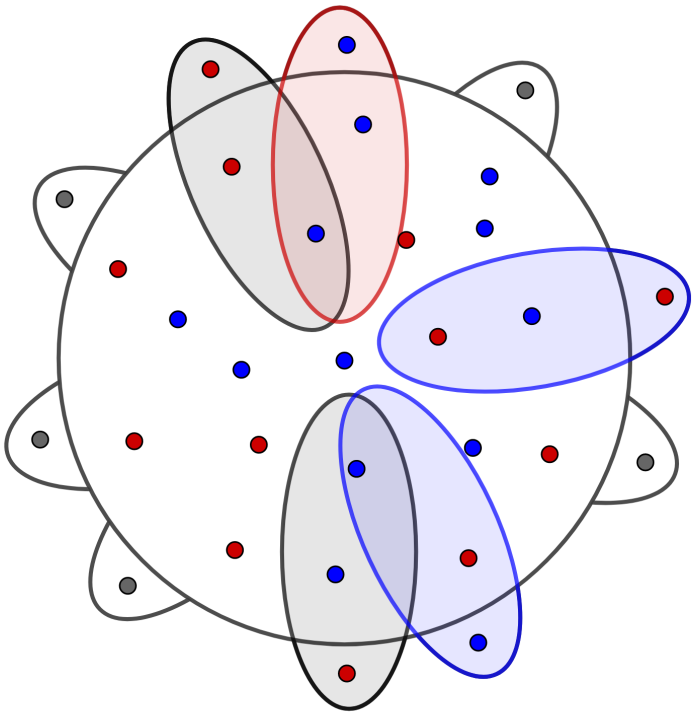
Global decoder G: construction

$\mathcal{D}_1$, assignment 2



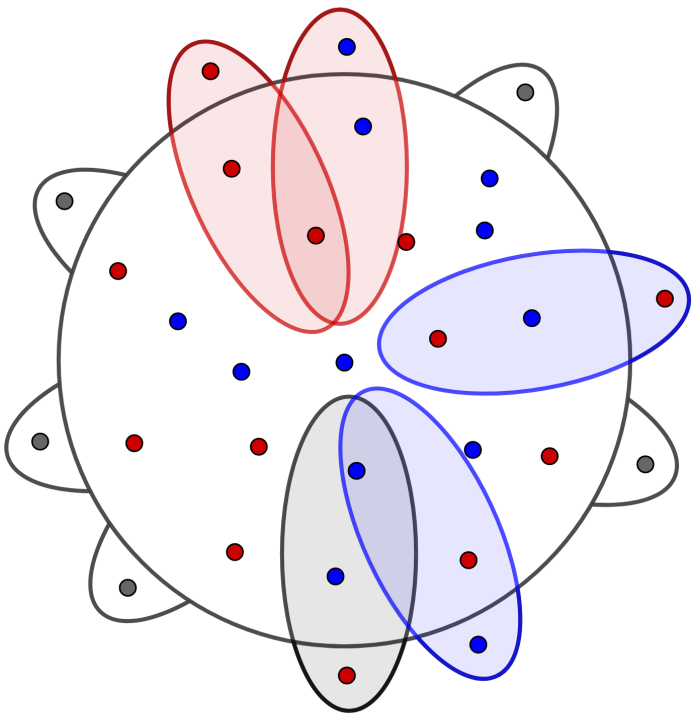
Global decoder G: construction

$\mathcal{D}_1$, assignment 3



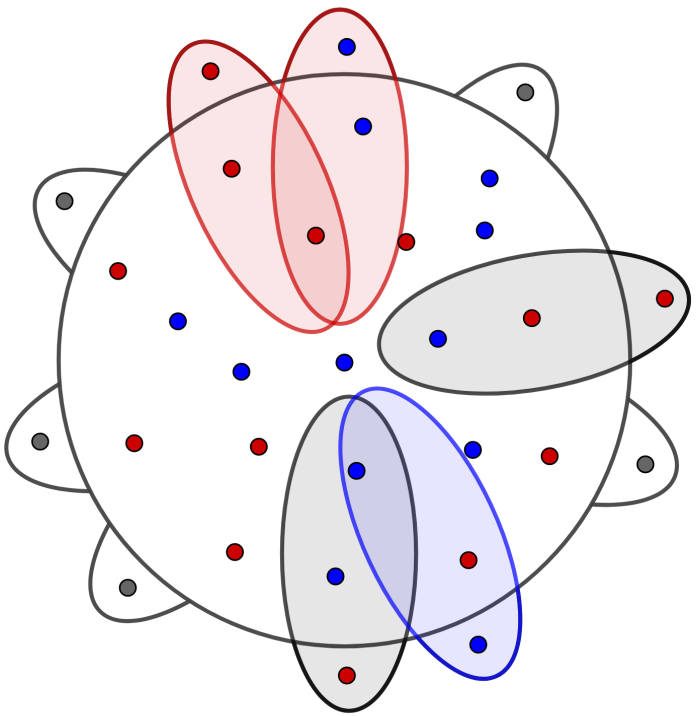
Global decoder G: construction

$\mathcal{D}_1$, assignment 4



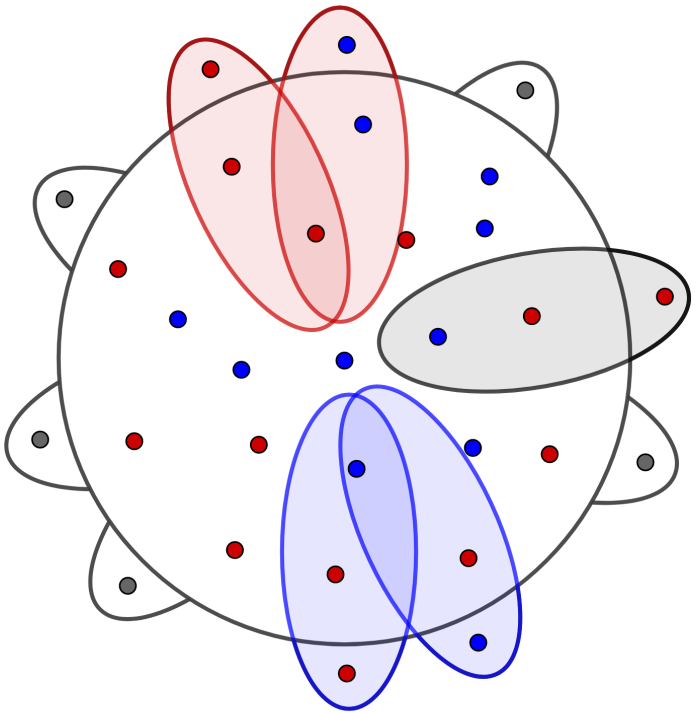
Global decoder G: construction

$\mathcal{D}_1$, assignment 5



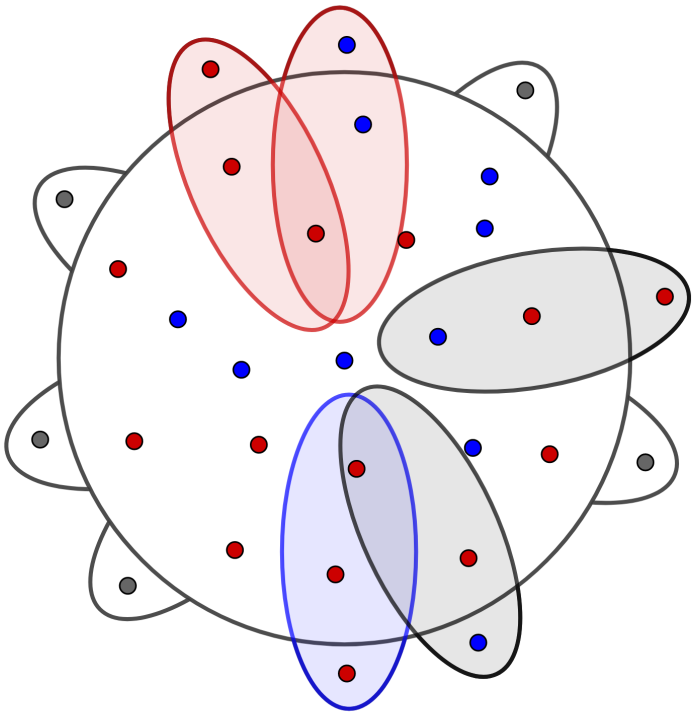
Global decoder G: construction

$\mathcal{D}_1$, assignment 6



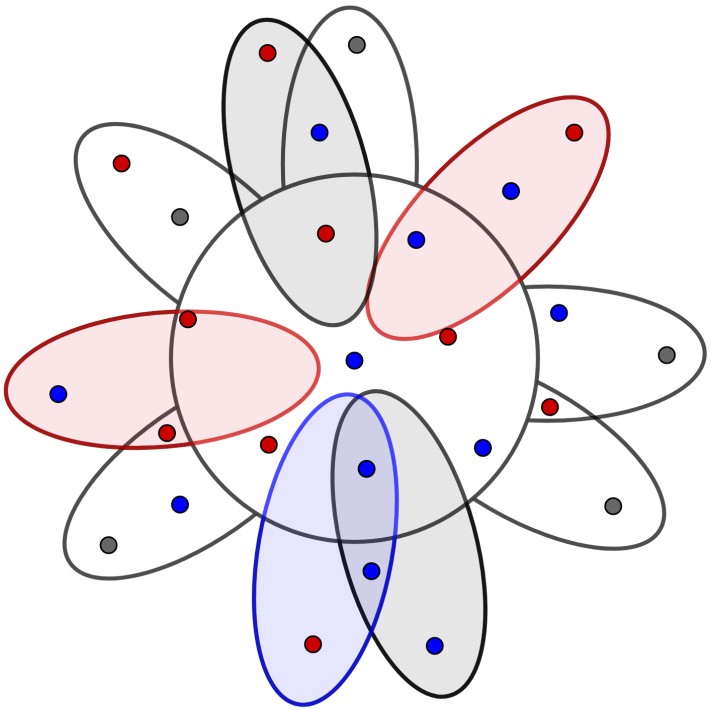
Global decoder G: construction

$\mathcal{D}_1$, assignment $2^{|K_1|}$



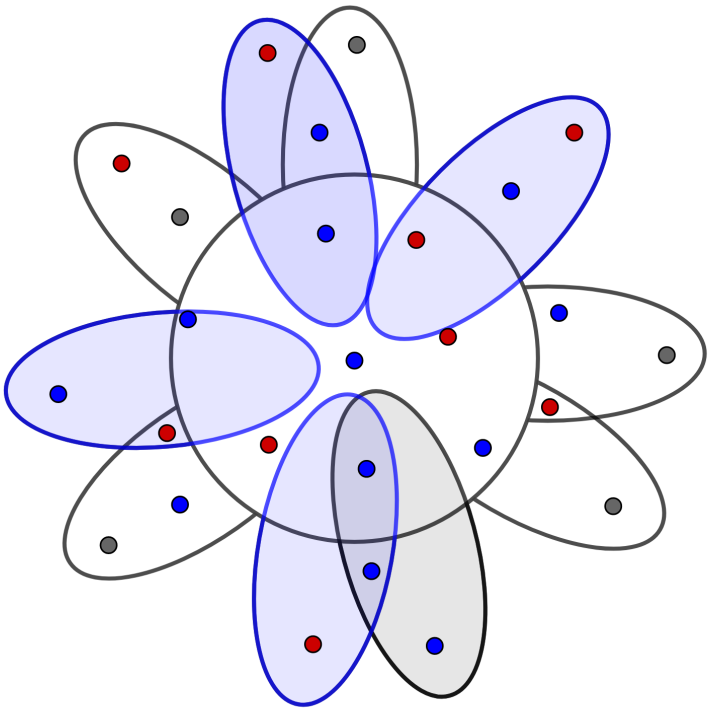
Global decoder G: construction

$\mathcal{D}_2$, assignment 1



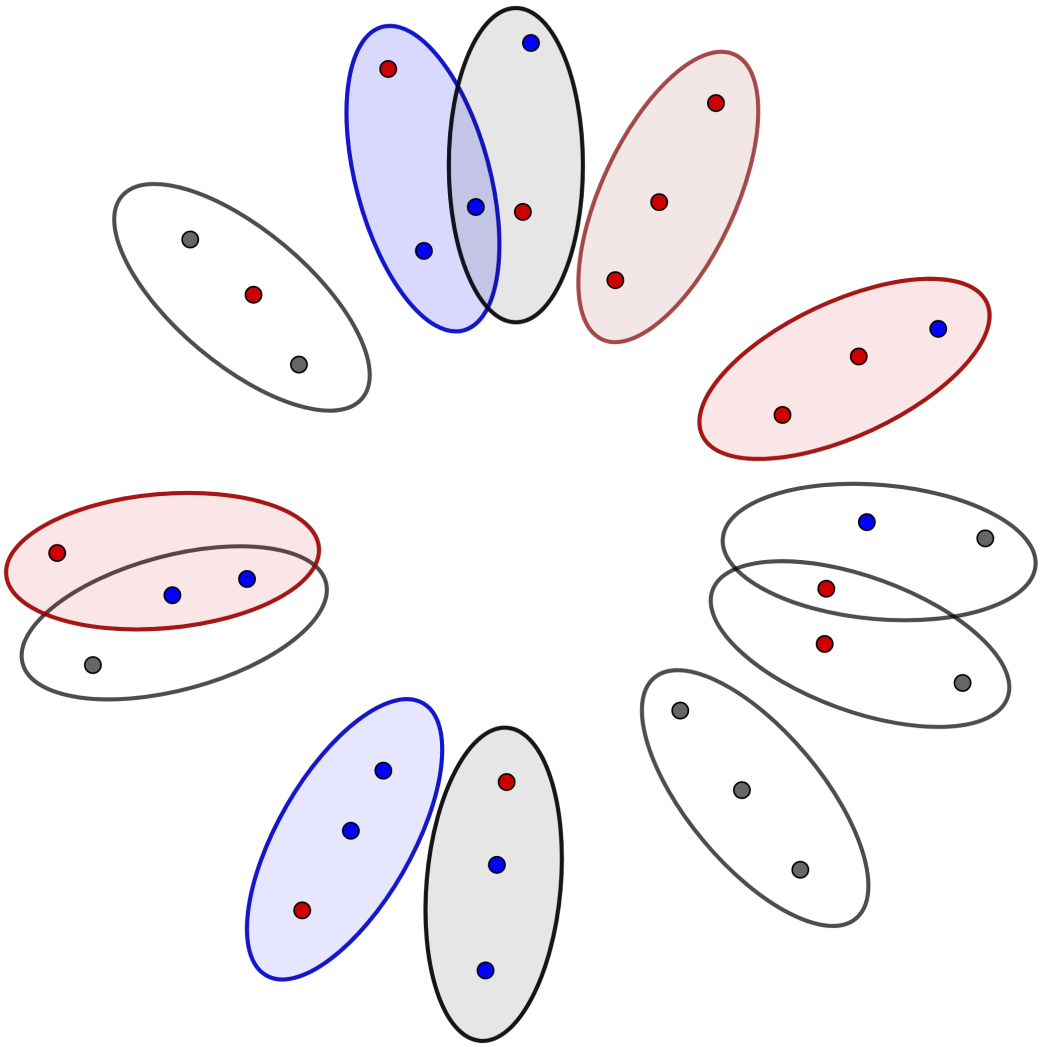
Global decoder G: construction

$\mathcal{D}_2$, assignment $\leq 2^{|K_2|}$: output •



Global decoder G: construction

$\mathcal{D}_3$, assignment $\leq 2^{|K_3|}$



Global decoder G: analysis

Volume lemma, upper bound

The *bad* q-sets (that decode to the wrong value) cover a small fraction of the codeword. Thus, their number is $O(n)$.

Global decoder G: analysis

Volume lemma, upper bound

The *bad* q-sets (that decode to the wrong value) cover a small fraction of the codeword. Thus, their number is $O(n)$.

Lemma (soundness)

For every daisy $\mathcal{D}_j$ and kernel assignment, the collection of bad queried q-sets is $< \tau_j$ with high probability.

Global decoder G: analysis

Volume lemma, lower bound

For some daisy $\mathcal{D}_j$, the queried q-sets cover a large fraction of the codeword. Thus, their number is $\Omega(n)$.

Global decoder G: analysis

Volume lemma, lower bound

For some daisy $\mathcal{D}_j$, the queried q-sets cover a large fraction of the codeword. Thus, their number is $\Omega(n)$.

Lemma (completeness)

For the correct guess, some daisy has a number of queried q-sets is at least $2\tau_j$ (so there are $\geq \tau_j$ *good* sets) with high probability.

Theorem [BGH+04]

There exist RLDCs with message length k and blocklength n satisfying

$$n = k^{1+O\left(\frac{1}{\sqrt{q}}\right)}.$$

Open problem

What is the largest $\alpha \in [1/2, 2]$ such that there exist RLDCs with

$$n = k^{1+\Omega\left(\frac{1}{q^{\alpha}}\right)}?$$

References


Ryan Alweiss, Shachar Lovett, Kewen Wu, and Jiapeng Zhang.

Improved bounds for the sunflower lemma, 2019.


Eli Ben-Sasson, Oded Goldreich, Prahladh Harsha, Madhu Sudan, and Salil P. Vadhan.

Robust PCPs of proximity, shorter PCPs and applications to coding.

In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing, Chicago, IL, USA, June 13-16, 2004*, pages 1–10, 2004.


P. Erdös and R. Rado.

Intersection theorems for systems of sets.

Journal of the London Mathematical Society, s1-35(1):85–90, 1960.


Tom Gur and Oded Lachish.

A lower bound for relaxed locally decodable codes.

CoRR, abs/1904.08112, 2019.