## 1. K-core

Let $G=(V,E)$ be a simple graph.

~~The~~ A (maximum) K-core is the maximum sized subgraph where all vertices have induced degree at least K. Equivalently, $S^*$ is a K-core if

- $\min_{v \in S^*} d_v(S^*) \geq K$
- If $T \subseteq V$ satisfies $\min_{v \in T} d_v(T) \geq K$ then $T \subseteq S^*$

* differs slightly from some definitions of K-core — we'll reconcile differences later.

Claim: $S^*$ is unique.

Assume not. Then assume $\tilde{S}$ is also a maximum K-core with $\tilde{S} \neq S^*$.

$Q = \tilde{S} \cup S^*$ is larger than $\tilde{S}$ and $S^*$ and contains both of them and $\min_{v \in Q} d_v(Q) \geq K$ and this contradicts the second point in the definition, hence $S^*$ must be unique.

## Algorithm (Peeling)

Consider an algorithm that repeatedly removes the min degree node.

Alg

$S_0 = V$

for $i = 1, 2, \cdots, n$

  $w = \text{argmin } d_v(S_{i-1})$

  if $d_w(S_{i-1}) \geq k$

    $\boxed{\text{return}}$ $S_{i-1}$

  else

    $S_i = S_{i-1} \setminus \{w\}$

  end

end

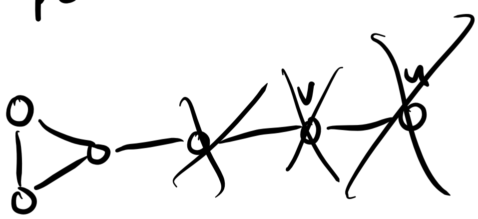
Let $T$ be the set returned

by this algorithm.

<u>Claim: $T = S^*$</u>

Observe that if a node is removed, it cannot be in the $K$-core, because it was removed at a point when it had induced degree $< K$.

Example: $K$-2




This means if $\boxed{v \notin T \text{ then } v \notin S^*}$

Next observe if $T$ is the set output by the algorithm, then for $j \in T$, $d_j(T) \geq K$ so $T$ certainly is contained in the $K$-core $S^*$ by the second point of the definition of $K$-core. In other words

if $\boxed{v \in T \text{ then } v \in S^*}$

So $T = S^*$. $\blacksquare$

Running time $\underline{O(E)}$

## Nested Structure

Above, we stopped as soon as we found the K-core, but if we keep going, we find all the K-cores.

<u>Def</u>: The "coreness" number of a node $v \in V$ is the value $c$ such that $v$ belongs to the $c$-core but not the $(c+1)$-core.

## Alg: All Cores

$n = |V|$

$S_0 = V$

currentcore = 0

coreness = zeros(n) // array of zeros

for $i = 1, 2, \ldots, n$

    $w = \text{argmin } d_v(S_{i-1})$

    $d_w = d_w(S_{i-1})$

    if $d_w >$ currentcore

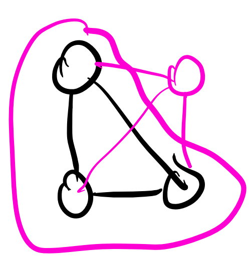
        currentcore $= d_w$

    end

    coreness[w] = currentcore

    $S_i = S_{i-1} \setminus \{w\}$

end



$K = 3$

**Proposition:** For any $K \in \mathbb{N}$ the K-core is given by one of

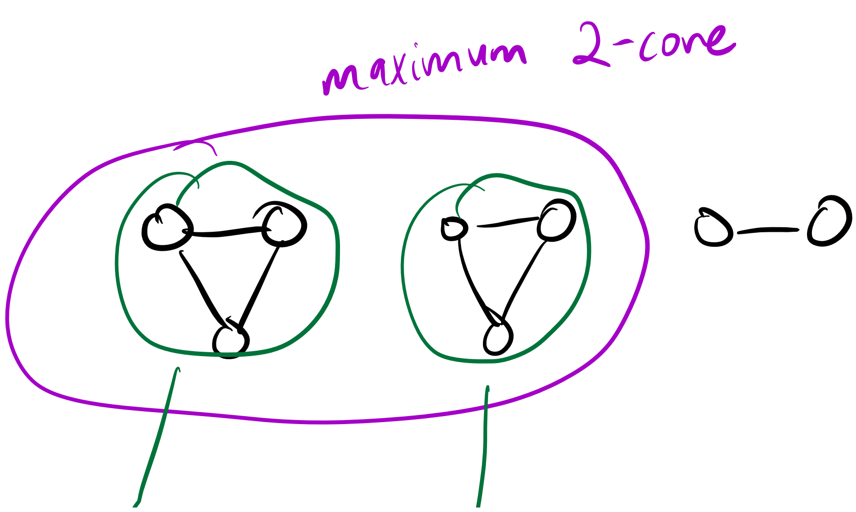
$$S_0 \supseteq S_1 \supseteq S_2 \supseteq \cdots \supseteq S_n = \phi$$

So the K-cores define a nested structure of subgraphs.

## Reconciling Definitions

Some definitions require a K-core to be a maximal connected subgraph where induced degree is $\geq K$ for all nodes.

E.g.



a 2-core a 2-core

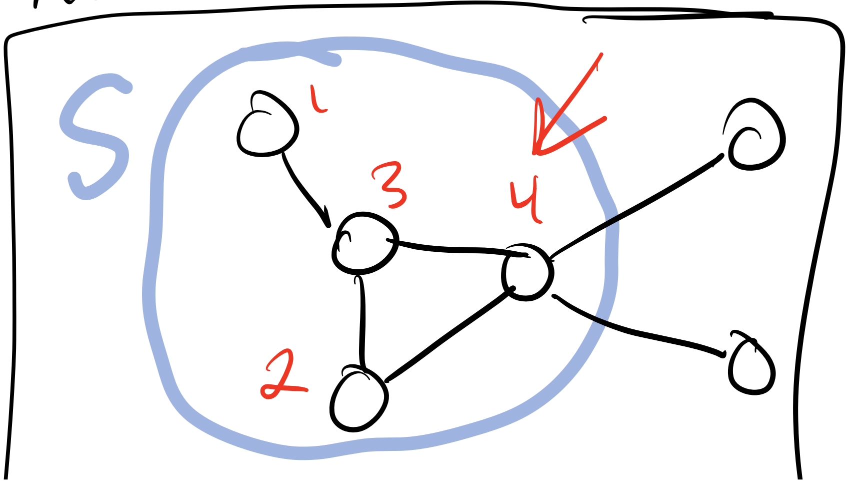
We can find these maximal connected K-cores by finding connected components of our maximum K-core.

---

## 2. Densest subgraph Problem

**Goal:** find $S^* = \underset{S \subseteq V}{\operatorname{argmax}} \frac{|E(S^*)|}{|S^*|}$

Recall that $\underline{vol(S)} - \underline{cut(S)} = \underline{2|E(S)|}$



$$\boxed{\mathrm{vol}(S) = 10}$$

$$|E(S)| = \frac{\mathrm{vol}(S) - \mathrm{cut}(S)}{2}$$

So we want to equivalently find

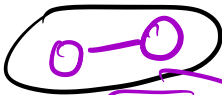
$$\boxed{\mathrm{argmax} = \frac{|E(S)|}{|S|}} = \mathrm{argmax}\ \frac{\mathrm{vol}(S) - \mathrm{cut}(S)}{2|S|}$$

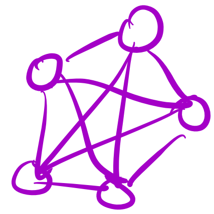
$$= \mathrm{argmax}\ \frac{\mathrm{vol}(S) - \mathrm{cut}(S)}{|S|}$$

$$= \boxed{\mathrm{argmin}\ \frac{|S|}{\mathrm{vol}(S) - \mathrm{cut}(S)}}$$

$$\frac{\frac{n(n-1)}{2}}{n}$$

Side note: $\gamma(S) = \frac{|E(S)|}{|S|}$





n nodes

Then $\boxed{\frac{1}{2}} \le \gamma(S^*) \le \boxed{\frac{n-1}{2}}$

(assuming here a connected simple graph with $|V| \geq 2$)

So $\frac{2}{n-1} \le \frac{2|S^*|}{vol(S^*) - cut(S^*)} \le 2$

$$\Rightarrow \frac{1}{n-1} \le \frac{|S^*|}{vol(S^*) - cut(S^*)} \le 1$$

$$f(S) = \frac{|S|}{vol(S) - cut(S)}$$

## $\boxed{\text{Decision Version}}$

For $\alpha \in \left[\frac{1}{n-1}, 1\right]$, is there some set $S \subseteq V$ such that $f(S) < \alpha$?

We can answer this by solving:

$$\boxed{\min_{S \subseteq V} g(S) = |S| - \alpha\, vol(S) + \alpha\, cut(S)}$$

Fact: $g(S) < 0$, then $f(S) < \alpha$

$$|S| - \alpha\, vol(S) + \alpha\, cut(S) < 0$$

$$|S| < \alpha\, vol(S) - \alpha\, cut(S)$$

$$f(S) = \frac{|S|}{vol(S) - cut(S)} < \alpha$$

Note that adding $\alpha \, vol(V)$ does not change the optimization:

$$\underset{S \subseteq V}{\arg\min} \; g(S) = \underset{S \subseteq V}{\arg\min} \; g(S) + \alpha \, vol(V)$$

$$= \underset{S \subseteq V}{\arg\min} \; |S| - \underbrace{\alpha \, vol(S) + \alpha \, vol(V)} + \alpha \, cut(S)$$

$$(1) = \boxed{\underset{S \subseteq V}{\arg\min} \; |S| + \alpha \, vol(\bar{S}) + \alpha \, cut(S)}$$

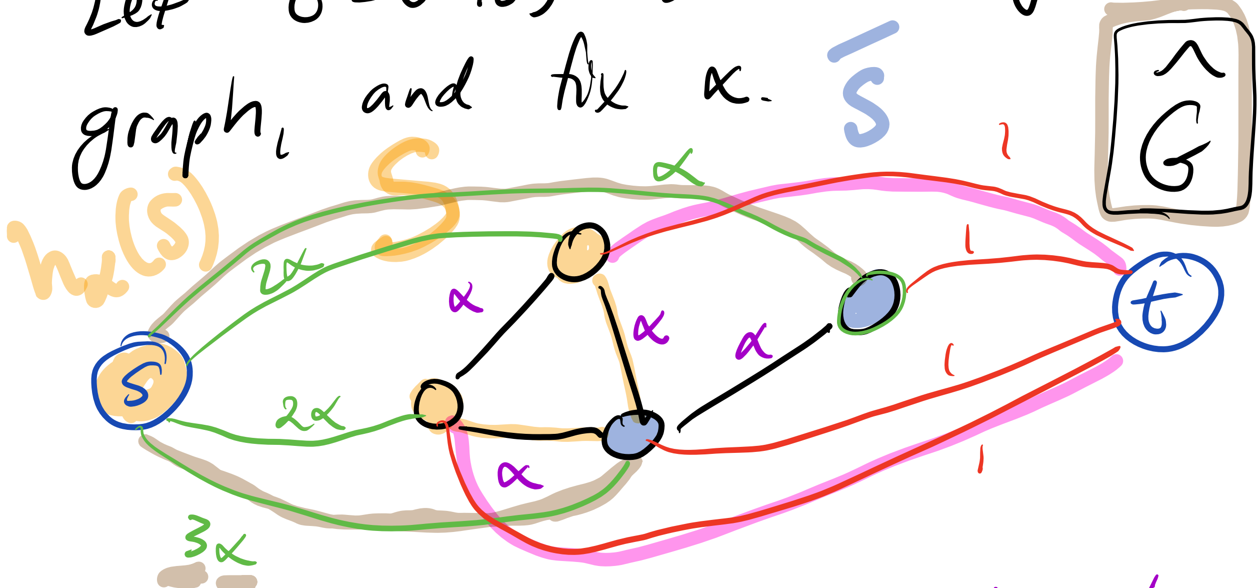
<u>Proposition</u>: Objective (1) can be solved via a single minimum s-t cut problem in an <u>augmented</u> graph.

$$\boxed{h_\alpha(S) = |S| + \alpha \, vol(\bar{S}) + \alpha \, cut(S)}$$

Note: $h_\alpha(S) < \alpha \, vol(V) \iff f(S) < \alpha$

# The s-t cut construction

Let $G=(V,E)$ be the original graph, and fix $\alpha$.




Step 1: weight all original edges by $\alpha$

Step 2: add (s) and (t) node

Step 3: For $v \in V$, add edge $(s,v)$ of weight $\alpha d_v$

Step 4: add edge $(v,t)$ of weight 1.

$$\boxed{h_\alpha(S)} = |S| + \alpha \operatorname{vol}(\bar{S}) + \alpha \operatorname{cut}(S)$$

**Proposition:** For every $S \subseteq V$ the value of the s-t cut $S \cup \{s\}$ in $\hat{G}$ is exactly $h_\alpha(S)$.

## Full Procedure for minimizing $f(S)$

- If $\min h_\alpha(S) < 0$, then $\exists$ a set $\hat{S}$ such that $f(\hat{S}) < \alpha$.

  We can then choose a new $\hat{\alpha} = f(\hat{S})$ and minimize $h_{\hat{\alpha}}(S)$.

- If $\min h_\alpha(S) = \alpha \, vol(V)$, then there is no $S$ with $f(S) < \alpha$.

$\boxed{\text{Algorithm}}$

$\alpha \leftarrow 1$

$S_\alpha = \text{argmin } h_\alpha(S)$

$\alpha_{new} = \underline{f(S_\alpha)} = \boxed{\dfrac{|S_\alpha|}{vol(S_\alpha) - cut(S_\alpha)}}$

while $\alpha_{new} < \alpha$

$\hat{S} = S_\alpha$

$\alpha = \alpha_{new}$

$S_\alpha = \text{argmin } h_\alpha(S)$

$\alpha_{new} = f(S_\alpha)$

end

return $\hat{S}$, $\gamma(\hat{S}) = \boxed{\dfrac{|E(\hat{S})|}{|\hat{S}|}}$

At most $2^n$ possible $\alpha$ values

$$\log 2^n \to O(n)$$

## 2. Peeling Algorithms

Recall that an equivalent problem is maximizing the average induced degree

$$\max_{S \subseteq V} \boxed{\delta(S) = \frac{\sum_{v \in S} d_v(S)}{|S|}} = \max_{S \subseteq V} \frac{2|E(S)|}{|S|}$$

A greedy peeling algorithm gives a 2-approximation.

**Alg**

$S_0 = V$

for $i = 1, 2, \ldots, n-1$

$\quad w_i = \underset{v \in S_{i-1}}{\operatorname{argmin}} \; d_v(S_{i-1})$

$\quad S_i = S_{i-1} \setminus \{w_i\}$

end

return $S_i$ with max average degree

**Proposition**: There is some $\underline{i}$ such that

$$\delta(S_i) \geq \frac{1}{2} \max_{S \subseteq V} \delta(S)$$

**Proof**: Let $d^* = \max \delta(S)$

$$= \delta(S^*) = \frac{\sum_{v \in S^*} d_v(S^*)}{|S^*|}$$

This means $\boxed{\sum_{v \in S^*} d_v(S^*) - d^* |S^*| = 0}$

Since $S^*$ is optimal, removing a single vertex would make the average degree worse, so for every $j \in S^*$

$$\frac{\left[\sum_{v \in S^*} d_v(S^*)\right] - 2d_j(S^*)}{\boxed{|S^*| - 1}} \leq d^*$$

$$\Rightarrow \cancel{\sum_{v \in S^*} d_v(S^*)} - 2d_j(S^*) \leq \cancel{d^*|S^*|} - d^*$$

$$\Rightarrow \quad d_j(S^*) \geq \frac{d^*}{2}$$

The greedy algorithm removes one node at a time, inducing an order $w_1, w_2, \ldots, w_{n-1}$

Let $j$ be the first node in $S^*$ that I remove during this process. Consider $S_{j-1}$. Note that $S_{j-1} \supseteq S^*$ hence $d_v(S_{j-1}) \geq d_v(S^*)$ $\forall v \in S^*$

We know $d_j(S_{j-1})$ is the smallest induced degree in $S_{j-1}$, so it is smaller than the average degree in $S_{j-1}$:

$$\underline{\underline{\frac{d^*}{2}}} \leq d_j(S^*) \leq d_j(S_{j-1}) \leq \frac{\sum_{i \in S_{j-1}} d_i(S_{j-1})}{|S_{j-1}|} = \underline{\underline{\delta(S_{j-1})}}$$

QED.

## 3. LP for densest subgraph problem

$$\max \sum_{(i,j)\in E} x_{ij}$$

$$\boxed{f(S) = \frac{|E(S)|}{|S|}}$$

$$\forall\, ij \in E \quad x_{ij} \le y_i$$

$$\boxed{x_{ij} = \min\{y_i, y_j\}} \qquad x_{ij} \le y_j$$

$$\boxed{\sum_{i=1}^{n} y_i \le 1}$$

$$x_{ij}, y_i \ge 0$$

<u>Prop 1</u>: Let OPT be the optimal LP solution. For every $S \subseteq V$ $f(S) \le OPT$.

Proof: Let $x = \frac{1}{|S|}$, and set

$$\hat{y}_i = x \quad \forall \quad i \in S$$

$$\hat{x}_{ij} = x \quad \forall \quad ij \in E(S)$$

Set all other variables to zero. Then

$$\sum_{ij \in E} x_{ij} = \sum_{ij \in E(S)} \hat{x}_{ij} + 0 = \sum_{ij \in E(S)} \frac{1}{|S|} = \boxed{\frac{|E(S)|}{|S|}}$$

This is a feasible LP solution:

$ij \in E(S)$ $\quad \hat{x}_{ij} = x = y_i = y_j$

$ij \in \partial S$ $\quad \hat{x}_{ij} = 0 < x$

$$\sum_{i=1}^{n} y_i = \sum_{i \in S} \frac{1}{|S|} = 1$$

**Prop**: We can find a set $S$ such that $f(S) \geq OPT$.

**Proof**: Let $\{\underline{x}, \underline{y}\}$ be the optimal LP solution.

Assume WLOG $x_{ij} = \min\{y_i, y_j\}$

For a parameter $r \geq 0$ define

$$S_r = \{ i \in V : \underline{y_i \geq r} \}$$

$$E_r = \{ (i,j) \in E : \underline{x_{ij} \geq r} \}$$

Observe $\underline{E_r} = \underline{E(S_r)}$

If $\underline{(i,j)} \in E_r$, then $\left.\begin{array}{l} \boxed{y_i} \geq x_{ij} \geq r \\ \boxed{y_j} \geq x_{ij} \geq r \end{array}\right\} \begin{array}{l} i \in S_r \\ j \in S_r \end{array}$

$$\Rightarrow ij \in \boxed{E(S_r)}$$

If $ij \in E(S_r)$ this $i \in S_r$, $j \in S_r$

so $y_i \geq r$, $y_j \geq r$

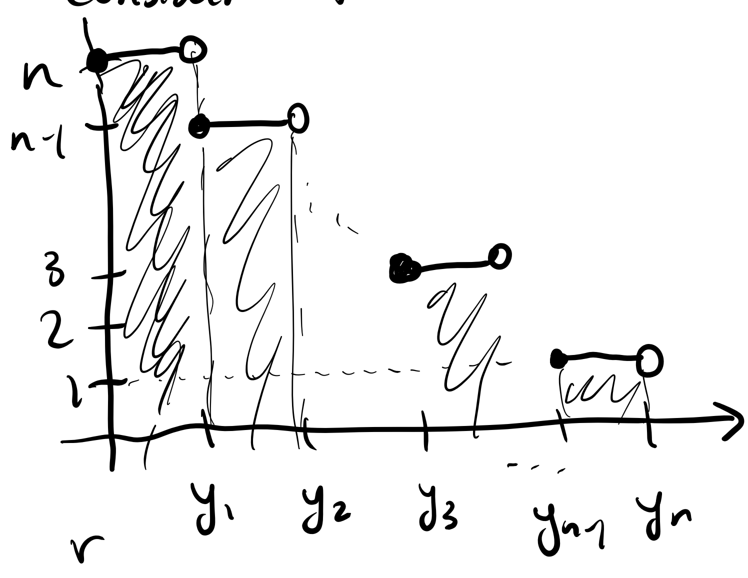
so $x_{ij} = \min\{y_i, y_j\} \geq r$

$\Rightarrow ij \in E_r$.

Order nodes so that

$$y_1 \leq y_2 \leq \cdots \leq y_n$$

Consider the function $g(r) = \boxed{|S_r|}$




$$\int_0^\infty |S_r|\, dr = \sum_{i=1}^n y_i$$

$$\int_0^\infty |S_r|\, dr = n\cdot(y_1 - 0) + (n-1)(y_2 - y_1)$$

$$+ (n-2)(y_3 - y_2) \cdots\cdots 2\cdot(y_{n-1} - y_{n-2}) + 1\cdot(y_n - y_{n-1})$$

$$= y_1 + y_2 + \cdots + y_n = \sum_{i=1}^{n} y_i \leq 1$$

Similarly: $$\int_0^\infty |E_r|\, dr = \sum_{ij \in E} x_{ij} = OPT$$

Claim: There is some $r \geq 0$ such that

$$f(S_r) = \frac{|E_r|}{|S_r|} \geq OPT$$

If this were not true, then

$$\int_0^\infty |E(r)|\, dr < \int_0^\infty |S_r| \cdot OPT\, dr$$

$$\Rightarrow \quad \sum_{ij \in E} x_{ij} = OPT < OPT \cdot 1$$

$\exists$ some $S_r$ $f(S_r) \geq OPT$.