

# Testing Graph Properties with Container Method

FOCS - 23

$G$  is  $\epsilon$ -far from property  $P$

if add/remove at least  $\epsilon n^2$  edges can get a  $G'$  with  $P$ .

$T$  is a canonical  $\epsilon$ -tester with sample cost  $s$ .

if  $T$  is a bounded-error randomized algorithm that samples a set  $S$  of  $s$  vertices and only examine  $G[S]$  to distinguish.

$G$  with  $P$  from  $G$   $\epsilon$ -far from  $P$ .

Main Conclusion:

$P$ -Clique /  $P$ -IndepSet: whether  $P$  has  $\text{PHR}$  clique/indset

$$S_{P\text{-IndepSet}}(n, \epsilon) = O\left(\frac{\rho^3}{\epsilon^2} \ln^3\left(\frac{1}{\epsilon}\right)\right)$$

$k$ -colourable:

$$S_{k\text{-colorable}}(n, \epsilon) = O\left(\frac{k}{\epsilon} \ln^2\left(\frac{1}{\epsilon}\right)\right)$$

In fact, Feige, Langberg and Schechtman ~~prove~~ prove

$$S_{P\text{-IndepSet}}(n, \epsilon) = \tilde{O}\left(\frac{\rho^3}{\epsilon^2}\right)$$



Container Method :Algorithm:Input:  $G$ , independent set  $I$ 1.  $F_0 \leftarrow \emptyset$ ,  $C_0 \leftarrow V$ 2. for  $t = 1, 2, \dots, |I|$  do3.  $v_t \leftarrow$  the vertex in  $I \setminus F_{t-1}$  with largest degree  
in  $G[C_{t-1}]$ ,4.  $F_t \leftarrow F_{t-1} \cup \{v_t\}$ 5.  $C_t \leftarrow C_{t-1} \setminus \{w \in C_{t-1} : (v_t, w) \in E\}$ // remove all neighbours of  $v_t$ 6.  $G_t \leftarrow G_t \setminus \{w \in G_{t-1} : \deg_{G[G_t]}(w) > \deg_{G[C_t]}(v_t)\}$ 7. Return  $F_1, \dots, F_{|I|}, C_1, \dots, C_{|I|}$ .

fingerprint container

Properties:①  $F_1 \subseteq F_2 \subseteq \dots \subseteq F_{|I|} = I \subseteq C_1 \subseteq \dots \subseteq C_t \subseteq C_{|I|}$ ② If for some  $t$ ,  $F_t(I_1) = F_t(I_2)$ .then,  $C_t(I_1) = C_t(I_2)$ ③  $\forall I, t \geq 1, \Delta(G[C_t(I)]) \leq \frac{n}{t}$ .Observation: ①  $\forall i, \Delta(G[C_i]) \leq \deg v_i$   
②  $\Delta(G[C_i]) \downarrow$  with  $i \leq |C_{t-1}| - |C_t|$ 

$\rho$ -IndepSet :

Thm (Graph Container Lemma for  $\rho$ -IndepSet)

Let  $G$  be  $\varepsilon$ -fair from  $\rho$ -IndepSet.

If  $I \in \mathcal{I}(G)$ , there exists an index  $t \leq \frac{\rho^2}{\varepsilon} \ln(\frac{2\rho}{\varepsilon})$

$$\text{s.t. } |C_t| \leq (\rho - t \cdot \frac{\varepsilon}{\rho \ln(\frac{2\rho}{\varepsilon})})n$$

and  $G[C_t]$  contains  $\leq \frac{\varepsilon n^2}{4}$  edges.

Thm  $S_{\rho\text{-IndepSet}}(n, \varepsilon) = O\left(\frac{\rho^3}{\varepsilon^2} \ln^3\left(\frac{1}{\varepsilon}\right)\right)$

Pf. Let  $S = C \cdot \frac{\rho^3}{\varepsilon^2} \ln^3\left(\frac{1}{\varepsilon}\right)$

Uniformly at random sample  $S = \{u_1, \dots, u_S\}$ .

Test  $S$  whether  $G[S]$  has  $f_S$ -size indepset

① If  $G$  has a  $f_n$ -size indepset  $I$

$X := \#$  nodes in  $I$  that  $\in S$  ~ Hypergeometric Distribution

$$P[S \text{ contains } f_S \text{ of } I] \geq \frac{1}{2}$$

② If  $G$  is  $\varepsilon$ -fair from  $\rho$ -IndepSet

Goal:  $P[S \text{ contains a } f_S \text{-size indepset}] \geq \frac{1}{2}$

By Graph Container Lemma for  $\rho$ -IndepSet,

If  $I \in \mathcal{I}(G)$ ,  $\exists t \leq \frac{\rho^2}{\varepsilon} \ln\left(\frac{2\rho}{\varepsilon}\right)$  s.t.

$$|C_t(I)| \leq (\dots)_D n$$



If  $S$  contains a  $(s-t)$ -size independent set  $I$ ,  
then  $\exists t \leq \frac{8p^2 \ln(\frac{2}{\delta})}{\varepsilon}$  s.t.

$$F_t \subset I \subset G_t \text{ and } G_t \leq D_n$$

$S$  contains a  $t$ -size  $F_t$ .

once  $F_t$  is determined,  $G_t$  is fixed.

then ~~the rest~~  $p_{s-t}$  nodes of  $I \setminus S \setminus F_t$   
should be chosen from  $G_t$ .

Let  $X$  be # nodes of  $G_t$  chosen by  $S \setminus F_t$ .

$$X \sim H(n, \frac{s-t}{D_n})$$

$$\mathbb{E}X = D(s-t)$$

$$\begin{aligned} \Pr[X \geq p_{s-t}] &\stackrel{\text{Chernoff bound.}}{\leq} \exp\left(-\frac{(p_{s-t} - \mathbb{E}X)^2}{p_{s-t} + \mathbb{E}X}\right) \\ &\leq \exp\left(-\frac{t^2 \sum s}{512 p^3 \ln^2\left(\frac{2}{\delta}\right)}\right) \end{aligned}$$

Now union bound over all choices of  $F_t$ ,

$$\Pr[S \text{ contains } p_{s-t} \text{-size independent set}] \leq \sum_{t \leq \frac{8p^2 \ln(\frac{2}{\delta})}{\varepsilon}} \binom{s}{t} \exp\left(-\frac{t^2 \sum s}{512 p^3 \ln^2\left(\frac{2}{\delta}\right)}\right) < \frac{1}{3} \quad \square$$

Rmk:  $e(G[G_t]) \leq \varepsilon n^2$  is not used.



Now, we prove Graph Container Lemma for  $\rho$ -IndSet

Pf. Let  $C$  be the first container in the

sequence  $C_1, C_2, \dots, (C_{|I|+1}, \dots = I)$  that contains at most  $\frac{\varepsilon}{4} n^2$  edges.  $\stackrel{\text{define}}{=}$

Its existence is guaranteed since  $C_{|I|+1} = I$  has no edge.

Suppose  $|C| = (\rho - \alpha)n$

Since  $G$  is  $\varepsilon$ -far from  $\rho$ -IndSet,  $\frac{\varepsilon}{2\rho} \leq \alpha \leq \rho$ .

In the process of the container  $C_t$  getting smaller and smaller until reaching  $C$ , there're two stages.

1°  $|C_t|$  is still large. Intuitively,  $(C_t \setminus C)$  has many edges.

2°  $|C_t|$  is already small, however still has  $\geq \frac{\varepsilon}{4} n^2$  edges.

Then intuitively, average degree is large.

By our greedy algorithm,  $G_t$  will soon lose its edges and reach  $C$ .

We first prove some technical lemmas.

Lemma 10:  $G$  has average degree  $2d$ , then at least  $d$  vertices have degree  $d$ .  
Suppose

Pf.

$$\frac{x(x-1)}{2} + e + \frac{d(n-x)-e}{2} > nd$$

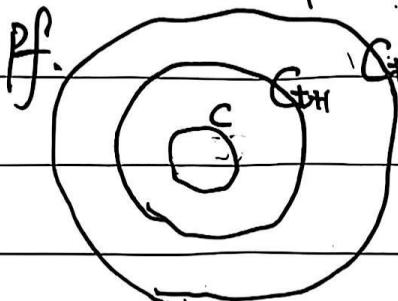
$$e \leq x(n-x)$$

We have.  
 $x > d$

to 7.b. Lemma 2: Let  $G$  be  $\varepsilon$ -far from  $\rho$ -Independent Set.

$I$  is an independent set.  $t \leq |I|$  is an index s.t.  $|G_t| \geq p_n$ . Suppose  $C \subseteq G_{t+1}$  has  $(\rho - \alpha)n$  vertices and  $G[C]$  contains  $\leq \frac{\varepsilon}{4\rho\alpha} n^2$  edges.

then  $|G_{t+1} \setminus C| \leq \left(1 - \frac{\varepsilon}{4\rho\alpha}\right) |G_t| t$



If  $C$  contains a vertex  $v$  with degree  $\geq \frac{\varepsilon}{4\rho\alpha} |G_t| t$  in  $G[G_t]$ , then the vertex  $v_{t+1}$  selected by the ~~greedy~~ algorithm must have ~~degree~~ at least  $\frac{\varepsilon}{4\rho\alpha} |G_t| t$ .

$$|G_{t+1} \setminus C| \leq \left(1 - \frac{\varepsilon}{4\rho\alpha}\right) |G_t| t$$

If  $C$  has no vertex with degree  $\geq \frac{\varepsilon}{4\rho\alpha} |G_t| t$  in  $G[G_t]$ ,

~~Now our goal is still to prove  $\deg v_{t+1} \geq \frac{\varepsilon}{4\rho\alpha} |G_t| t$~~   
~~We use the last lemma, aiming to prove~~  
 ~~$e(G[G_t])$~~

We claim:  $G[G_t \setminus C]$  has  $\geq \frac{\varepsilon}{4\rho\alpha} |G_t| t$  edges.

If we admit this claim,

by the last lemma, at least  $\frac{\varepsilon}{4\rho\alpha} |G_t| t$  vertices in  $G[G_t \setminus C]$  has degree  $\geq \frac{\varepsilon}{4\rho\alpha} |G_t| t$



Now, if  $\deg v_{t+1} \geq \frac{\epsilon}{4\alpha} |G_t \setminus C|$ , then  $N(v_{t+1})$  will be excluded by  $C_{t+1}$ .

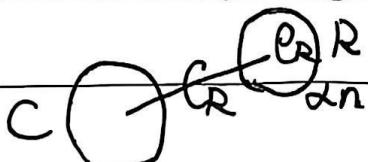
If  $\deg v_{t+1} < \frac{\epsilon}{4\alpha} |G_t \setminus C|$ , then those vertices with degree  $\geq \frac{\epsilon}{4\alpha} |G_t \setminus C|$  will be excluded.

$$\text{So } |C_{t+1} \setminus C| \leq \left(1 - \frac{\epsilon}{4\alpha}\right) |G_t \setminus C|$$

Now we prove the claim.

The idea is since  $G[C]$  contains  $\approx \frac{\epsilon}{4} n^2$  edges and  $G$  is  $\epsilon$ -far from  $\ell$ -IndepSet, all  $\alpha n$ -size subgraph of  $G[G_t \setminus C]$  must have many edges.

The original paper adopted a fancy probabilistic method, but here we use an equivalent counting.



We denote  $C_R$  as # edges across  $C, R$

$e_R$  as # edges in  $R$

$$(p\alpha)n. \quad \text{If } R, \quad C_R + e_R \geq \frac{3}{4} \epsilon n^2$$

$$\sum_{\text{all } \alpha n\text{-size } R} (C_R + e_R) \geq \frac{3}{4} \epsilon n^2 \left(\frac{m}{\alpha n}\right)$$

where  $m$  denotes  $|G_t \setminus C|$

$$\binom{m-1}{\alpha n-1} \left( \sum_{v \in C} \deg v - 2e(C) \right) + \binom{m-2}{\alpha n-2} e(G \setminus G \setminus C)$$

$$\text{If } v \in C, \deg v < \frac{\epsilon}{4\alpha} m \geq \frac{3}{4} \epsilon n^2 \left(\frac{m}{\alpha n}\right)$$

$$e(G \setminus G \setminus C) \geq \frac{1}{2} \epsilon n \frac{m(m-1)}{\alpha(\alpha n-1)} \geq \frac{\epsilon}{4\alpha} m^2. \quad \square$$

①



CS 扫描全能王

3亿人都在用的扫描App

Now back to ~~Container~~ proof of container lemma.

Stage 1°:  $|C_t| \geq p_n$ .

Suppose  $t^*$  be the largest index s.t.  $|C_{t^*}| \geq p_n$ .

Then  $|C_{t^*} \setminus C| \leq \left(1 - \frac{\varepsilon}{4\rho\alpha}\right)^{t^*} \cdot n$ .

by Lemma 2.

$$\begin{aligned} \text{Since } |C_{t^*} \setminus C| &\geq \alpha n, \quad t^* \leq \frac{4\rho\alpha}{\varepsilon} \ln\left(\frac{1}{\alpha}\right) \\ &\leq \frac{4\rho\alpha}{\varepsilon} \ln\left(\frac{2\rho}{\varepsilon}\right) \end{aligned}$$

Stage 2°:  $|C_t| < p_n$  but  $G[C_t]$  contains  $\geq \frac{\varepsilon}{4} n^2$  edges.

then by lemma 1,

at least  $\frac{\varepsilon}{4\rho} n$  vertices has degree  $\geq \frac{\varepsilon}{1 + \rho} n$  in  $G[C_t]$

Similarly, at least  $\frac{\varepsilon}{4\rho} n$  vertices will be removed in the  $t$ -th iteration.

Since  $|C_{t+1} \setminus C| \leq \alpha n$

this stage will last  $\frac{4\rho\alpha}{\varepsilon}$  iterations.

Summing it up, there're  $\leq \frac{8\rho^2}{\varepsilon} \ln\left(\frac{2\rho}{\varepsilon}\right)$  iterations before we reach  $C$   $\square$ .

