Testing Graph Properties with Container Method

FOCS '23

G is ε-far from propery P
if add/remove at ~~least~~ least εn^2 edges can get a G' with P.

T is a canonical ε-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 ε-far from P.

Main Conclusion:

ρ-Clique / ρ-IndepSet : whether has ρ|V| clique/indset

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

k-colourable:

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

In fact, Feige, Langberg and Schechtman 2004 ~~prove~~ prove

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

Date. ________ NO. ________

Container Method:

Algorithm:

Input: G, independent set I

1. $F_0 \leftarrow \phi$. $C_0 \leftarrow V$
2. for $t = 1, 2, \ldots, |I|$ do
3. $\quad v_t \leftarrow$ the vertex in $I \setminus F_{t-1}$ with largest degree in $G[C_{t-1}]$;
4. $\quad F_t \leftarrow F_{t-1} \cup \{v_t\}$
5. $\quad C_t \leftarrow C_{t-1} \setminus \{w \in C_{t-1} : (v_t, w) \in E\}$ // remove all neighbours of v_t
6. $\quad C_t \leftarrow C_t \setminus \{w \in C_{t-1} : \deg_{G[C_{t-1}]}(w) > \deg_{G[C_{t-1}]}(v_t)\}$
7. Return $\underbrace{F_1, \ldots, F_{|I|}}_{\text{fingerprint}}$ $\underbrace{C_1, \ldots, C_{|I|}}_{\text{container}}$.

Properties:

① $F_1 \subseteq F_2 \subseteq \cdots \subseteq F_{|I|} = I \subseteq C_{|I|} \subseteq \cdots \subseteq C_2 \subseteq C_1$

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 $\forall i$ $\Delta[G[C_i]] \leq \deg v_i \leq |C_{i-1}| - |C_i|$

② $\Delta[G[C_i]] \downarrow$ with i

ρ-IndepSet:

Thm (Graph Container Lemma for ρ-IndepSet)

Let G be ε-far from ρ-IndepSet.

$\forall I \in \mathcal{I}(G)$, there exists an index $t \le \frac{8\rho^2}{\varepsilon}\ln\left(\frac{2\rho}{\varepsilon}\right)$

s.t. $|C_t| \le \left(\rho - t\cdot\frac{\varepsilon}{8\rho\ln\left(\frac{2\rho}{\varepsilon}\right)}\right)n$

and $G[C_t]$ contains $\le \frac{\varepsilon n^2}{4}$ 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 ρs-size indepset

① If G has a ρn-size indepset I

X: # nodes of I that S contains $\sim$ Hypergeom Distribution

$\mathbb{P}[S \text{ contains } \rho s \text{ of } I] \approx \frac{1}{2}$.

② If G is ε-far from ρ-IndepSet

Goal: $\mathbb{P}[S \text{ contains a } \rho s\text{-size indepset}]$

By Graph Container Lemma for ρ-IndepSet,

$\forall I \in \mathcal{I}(G)$, $\exists t \le \frac{8\rho^2\ln\left(\frac{2\rho}{\varepsilon}\right)}{\varepsilon}$ s.t.

$|C_t(I)| \le \left(\frac{\cdots}{D}\right)n$

If S contains a ρs-size indep set I,

then $\exists\ t \le \frac{8\rho^2 \ln(\frac{2\rho}{\varepsilon})}{\varepsilon}$ s.t.

$F_t \subset I \subset C_t$ and $C_t \le Dn$

S contains a t-size F_t.

once F_t is determined, C_t is fixed.

then ~~the rest~~ the rest $\rho s - t$ nodes of I ~~should~~ $\subset S \setminus F_t$ should be chosen from C_t.

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

$X \sim H(n, s-t, Dn)$

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

$$\mathbb{P}[X \ge \rho s - t] \overset{\text{Chernoff bound}}{\le} \exp\left(\frac{-(\rho s - t - \mathbb{E}X)^2}{\rho s - t + \mathbb{E}X}\right)$$

$$\le \exp\left(-\frac{t^2 \varepsilon^2 s}{512 \rho^3 \ln^2(\frac{2\rho}{\varepsilon})}\right)$$

Now union bound over all choices of F_t,

$$\mathbb{P}[S \text{ contains } \rho s\text{-size independent set}] \le \sum_{t \le \frac{8\rho^2 \ln(\frac{2\rho}{\varepsilon})}{\varepsilon}} \binom{s}{t} \exp\left(-\frac{t^2 \varepsilon^2 s}{512 \rho^3 \ln^2(\frac{2\rho}{\varepsilon})}\right) < \frac{1}{3} \quad \square$$

<u>Rmk</u>: $e(G[C_t]) \le \varepsilon n^2$ is not used.

Now, we prove Graph Container Lemma for ρ-IndSet

Pf. Let C be the first container in the sequence $C_1, C_2, \cdots$ (define $C_{|I|+1}, \cdots = I$) that contains at most $\frac{\varepsilon}{4} n^2$ edges.

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

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

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

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. $|C_t \setminus C|$ will soon shrink.

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

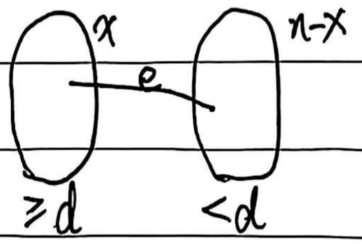
Then intuitively, average degree is large.

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

We first prove some technical lemmas.

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

Pf.



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

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

We have

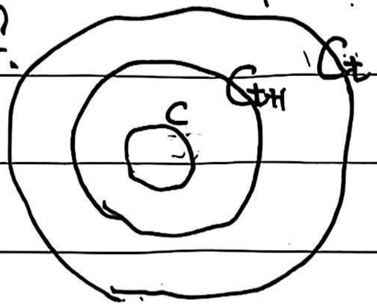
$$x > d$$

Date.__________ NO.__________

<u>Lemma 2</u>: Let G be ε-far from ρ-IndepSet.

I is an independent set. $t < |I|$ is an index s.t. $|C_t| \geq \rho n$. Suppose $C \subseteq C_{t+1}$ has $(\rho-\alpha)n$ vertices and $G[C]$ contains $\leq \frac{\varepsilon}{4} n^2$ edges, then $|C_{t+1} \setminus C| \leq \left(1 - \frac{\varepsilon}{4\rho\alpha}\right) |C_t \setminus C|$.

Pf.



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

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

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

~~now our goal is still to prove $\deg v_{t+1} \geq \frac{\varepsilon}{4\rho\alpha}|C_t \setminus C|$~~

~~We use the last lemma, aiming to prove $e(G[C_t])$~~

We claim: $G[C_t \setminus C]$ has $\geq \frac{\varepsilon}{4\rho\alpha}|C_t \setminus C|^2$ edges.

If we admit this claim,

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

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

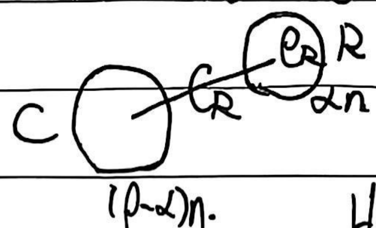
if $\deg v_{t+1} < \frac{\varepsilon}{4\rho\alpha}|C_t \setminus C|$, then those vertices with degree $\geq \frac{\varepsilon}{4\rho\alpha}|G_t \setminus C|$ will be excluded.

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

Now we prove the clam.

The idea is since $G[C]$ contains $\leq \frac{\varepsilon}{4}n^2$ edges and G is ε-far from ρ-IndepSet, all αn size subgraph of $G[C_t \setminus C]$ 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

$$\forall R,\ c_R + e_R \geq \frac{3}{4}\varepsilon n^2$$

$$\sum_{\text{all } \alpha n\text{-size } R} (c_R + e_R) \geq \frac{3}{4}\varepsilon n^2 \binom{m}{\alpha n}$$

where m denotes $|G_t \setminus C|$

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

$\forall v \in C$, $\deg v < \frac{\varepsilon}{4\rho\alpha} m$ $\qquad \geq \frac{3}{4}\varepsilon n^2 \binom{m}{\alpha n}$

$$e(G[G_t \setminus C]) \geq \frac{1}{2}\varepsilon n \frac{m(m-1)}{\alpha(\alpha n-1)} \geq \frac{\varepsilon}{4\rho\alpha} m^2. \ \square$$

Date. __________ NO. __________

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

Stage 1°: $|C_t| \geq \rho n$.

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

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

by Lemma 2.

Since $|C_{t^*} \setminus C| \geq \alpha n$, $t^* \leq \frac{4\rho\alpha}{\varepsilon} \ln\left(\frac{1}{\alpha}\right)$

$\leq \frac{4\rho\alpha}{\varepsilon} \ln\left(\frac{2\rho}{\varepsilon}\right)$

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

then by lemma 1,

at least $\frac{\varepsilon}{4\rho} n$ vertices has degree $\geq \frac{\varepsilon}{4\rho} n$ 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 at most 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 $\square$.

k-Colorability.

Thm (graph container lemma for k-colorability)

Let G be ε-far from k-colorable, and let $I_1, \dots, I_k$ be independent sets in G. Then, there exists $t \le \frac{4}{\varepsilon}$ s.t. $\left| \bigcup_{i=1}^{k} C_t(I_i) \right| \le \left(1 - t\frac{\varepsilon}{4\ln\frac{1}{\varepsilon}}\right)n$

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

Pf. Let $s = C\frac{k}{\varepsilon}\ln^2\frac{1}{\varepsilon}$ with constant C sufficiently large.

Sample $S = \{u_1, \dots, u_s\}$ u.a.r

If $G[S]$ is k-colorable $\Rightarrow$ Output YES

Else $\Rightarrow$ Output NO.

So If G is k-colorable, algorithm always accepts.

If G is ε-far from k-colorable but $G[S]$ is k-colorable.

Then $\exists$ independent sets $I_1, \dots, I_k$ s.t.

$$S = I_1 \sqcup \cdots \sqcup I_k$$

By graph container lemma for k-colorability,

$\exists t \le \frac{4}{\varepsilon}$ s.t. $\left| \bigcup_{i=1}^{k} C_t(I_i) \right| \le \left(1 - t\frac{\varepsilon}{4\ln\frac{1}{\varepsilon}}\right)n$

$$\begin{array}{ccccc} S = & I_1 & \sqcup \cdots \sqcup & I_k \\ & \cup & & \cup \\ & F_t(I_1) & & F_t(I_k) \end{array}$$

We union bound over all t-size sets $F_t(I_1), \dots, F_t(I_k)$

Once fix them, the rest $s-tk$ vertices of S must be chosen from $\bigcup_{i=1}^{k} C_t(I_i)$

The total probability is

$$\le \sum_{t=1}^{\frac{4}{\varepsilon}} \binom{0.S}{t}^{k} \left(1-\frac{t\varepsilon}{4\ln\frac{1}{\varepsilon}}\right)^{s-tk} < \frac{1}{3}$$

noting that only when $\varepsilon < \frac{1}{k}$ the problem is non-trivial

otherwise all graph is $\frac{1}{k}$-far from k-colorable.

Now we prove container lemma for k-colorability.

Pf. Towards a contradiction, assume

$$\forall t \le \frac{4}{\varepsilon},\ \left|\bigcup_{i=1}^{k} C_t(I_i)\right| > \left(1-\frac{t\varepsilon}{4\ln\frac{1}{\varepsilon}}\right)n$$

We'll construct a partition $V_1 \dots V_k$ of V with few edges inside $G[V_1] \dots G[V_k]$

Initialize $V_1 = C_{\frac{4}{\varepsilon}}(I_1) \dots V_k = C_{\frac{4}{\varepsilon}}(I_k)$

Remove vertices abitrarily until $V_1 \dots V_k$ forms a partition of $\bigcup_{i=1}^{k} C_{\frac{4}{\varepsilon}}(I_i)$

By Property ③, each vertex has degree $\le \frac{\varepsilon n}{4}$ in its set, so the sum of edges inside $G[V_1] \dots G[V_k] \le \frac{\varepsilon n^2}{4}$

Now we allocate the remaining vertices in $V \setminus \bigcup_{i=1}^{k} C_{\frac{4}{\varepsilon}}(I_i)$ to one of the partition while not increasing ~~two~~ too much edges.

For $t = 1, \cdots, \frac{4}{\varepsilon}$. let $A_t = \{ v : v \in \bigcup_{i=1}^{k} C_{t-1}(I_i)$ but $v \notin \bigcup_{i=1}^{k} C_t(I_i) \}$

Then $\bigsqcup_{t=1}^{\frac{4}{\varepsilon}} A_t = V \setminus \bigcup_{i=1}^{k} C_{\frac{4}{\varepsilon}}(I_i)$

For each vertex v in $V \setminus \bigcup_{i=1}^{k} C_{\frac{4}{\varepsilon}}(I_i)$,

suppose $v \in A_t$

then allocate v to V_i, where i is ~~the~~ an index that $v \in C_{t-1}(I_i)$

~~Thus~~ To count additional edges, consider ~~the~~ ordering t from $\frac{4}{\varepsilon}$ to 1, i.e. first allocate vertices in $A_{\frac{4}{\varepsilon}}$ down to A_1.

When $v \in A_t$ is allocated. to V_i,

$V_i \subseteq C_{t-1}(I_i)$

So ~~$\deg_{C_{t-1}(I_i)}$~~ $\deg_{G[V_i]} v \le \frac{n}{t-1}$

In total. all the edges inside $V_1 \cdots V_k$.

$$\le \frac{\varepsilon n^2}{4} + |A_1| \cdot n + \sum_{t=2}^{n} |A_t| \frac{n}{t-1}$$

Date. ________ NO. ________

$\cancel{\frac{\varepsilon n^2}{4}}$ subject to $\sum_{i=1}^{t} |A_i| \leq \frac{t\varepsilon n}{4\ln\frac{1}{\varepsilon}}$

$\forall\, t \leq \frac{4}{\varepsilon}$.

$$\# \text{edges} \leq \frac{\varepsilon n^2}{4} + \frac{\varepsilon n^2}{4\ln\frac{1}{\varepsilon}} + \frac{\varepsilon n^2}{4\ln\frac{1}{\varepsilon}} \sum_{t=2}^{4/\varepsilon} \frac{1}{t-1} < \varepsilon n^2$$

which is a contratridiction $\square$.